NOTICE: When government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related government procurement operation, the U. S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.

DEPARTMENT OF PSYCHOLOGY

KANSAS STATE UNIVERSITY

MANHATTAN

Technical Reports 5-8, 1962

The Evolution of Perceptual

Frames of Reference

These reports describe work conducted under contract Nonr-3634(01) between Kansas State University and the Physiological Psychology Branch, Office of Naval Research.

The principal investigator is William Bevan, Department of Psychology.

November, 1962

Technical Report No. 5, 1962

Patterns of Experience and the Perceived Orientation of the Necker Cube[1]

Edward D. Turner and William Bevan
Kansas State University

1. This report describes work under contract Nonr-3634(01) between Kansas State University and the Physiological Psychology Branch, Office of Naval Research. It is part of a project entitled, "The evolution of perceptual frames of reference."

This paper has been accepted for publication in the *Journal of General Psychology*.

November, 1962

A. Introduetion

Contrary to the traditional view concerning stimulus-response relationships in perceptual judgment, a rapidly growing literature has demonstrated that judgments of even the simplest sensory sort are the product of a complex of relationships and reflect, in part, the subject's experience with some range of stimuli within the class represented by the stimulus being judged. Given a constant judgmental background, the individual judgment depends upon tho statistical properties of the order of presentation within which it appears (5) and the relative impressiveness of the several stimuli on the occasions on which they have occurred. The impressiveness of a stimulus, in turn, depends upon certain relevant physical properties (3) as well as on operations attendant upon its presentation -- whether, for example, it is assigned some special role such as that of anchor (2). It is possible to predict with precision judgmental values on individual trials for certain traditionally-identified sensory dimensions as a differential between the present stimulus value and a weighted average of relevant stimuli previously experienced by the judge (adaptation level).

Recent studies in our laboratory have been directed toward examining perceptual dimensions in terms of the model of adaptation level. Anchor-induced shifts in judgment have been demonstrated for a shape dimension. Furthermore, for an anchor effect to oecur, it is clear that the anchor-designate, in addition to relating in some significant way to other stimuli on the property being judged, must share in common with these stimuli certain criterial attributes (1, 11). The perceptual property selected for experimental treatment is phenomenal ambiguity of orientation, a coneeptual rather than a sensory dimension (Conceptual as well as sensory dimensions have been shown to be amenable to treatment in terms of A.L. theory [4].), and the variables experimentally manipulated are average degree of ambiguity within the stimulus series, relative frequency of presentation for the several series members, and reinforcement of certain series members. Like any average, adaptation level is the joint product of individual stimulus values and their relative frequency of presentation. Reinforcement was included to explore the role of this factor in altering the impressiveness of the stimuli (the weight of their contribution to adaptation level) on individual trials.

B. Method

This study consists of three short experiments, one dealing with each of the three variables mentioned above. Essentially the same method was used in all. Individual subjects judged a series of brief presentations of the cube rendered to some degree or degrees unambiguous followed by several presentations of the ambiguous version as test trials. The same apparatus and instructions were used throughout.

1. Subjects. Thirty women students in an introductory psychology course served as subjects, ten in eaeh experiment. None were familiar with the purpose of the study.

2. Apparatus. Two unambiguous versions of the Necker Cube were prepared -- one invariably oriented to the left, the other to the right -- by shading certain faces and thickening certain contours. When presented in a mirror-type, two-field mixing tachistoscope they made possible serveral versions of the cube: completely unambiguous, left; moderately unambiguous, left; slightly unambiguous, left; completely ambiguous (the Necker Cube); slightly unambiguous, right; moderately unambiguous, right; completely unambiguous, right. This was accomplished by

---

Insert Figure 1 here

---

simultaneously increasing and decreasing the brightness of two fields in a complementary fashion, such that the total field brightness remained constant. Change was continuous from extreme left to extreme right. Adjustment of the brightness controls for the two fields allowed change in appearance from unambiguous left orientation through the balanced condition to unambiguous right orientation or vice versa.

3. General Procedure. Subjects were tested individually. After being shown how to view the stimulus field, each subject was told that she was participating in a study of her ability to discriminate fine differences. She was instructed that cubes would be briefly presented one after the other and that in each case she was to indicate the orientation, left or right, of the figure. She was then shown drawings of the left and right unambiguous versions. To further insure that she understood what she was to report upon, four to six presentations of each version were made tachistoscopically and she was required to report on orientation. Finally she received thirty trials for the purpose of establishing A.L., followed by three test trials to test the effect of the experimental treatment. Subjects who incorrectly identified orientation during the training trial were omitted prior to the test trials. Only two subjects were eliminated on these grounds. Thirty qualified for test.

C. Experiment I

1. Procedure. In Experiment I the variable subjected to manipulation was average degree of ambiguity of orientation. It was reasoned that if perceived orientation were conceptualized as a variable capable of continuous transformation from left to right or vice versa through changes in stimulus ambiguity, an order of presentation which resulted in the average ambiguity being lower for one orientation than for the other should result in the impression identified with lower average ambiguity becoming established as an orientational adaptation level. Accordingly, any stimuli with ambiguity scale values between this value and complete unambiguity of the opposite side should yield impressions of the opposite orientation. Thus, for example, if the right-oriented versions are, on the average, moderately unambiguous while those of the left are only slightly unambiguous, the A.L. should be identified with the right orientation and the balanced Necker Cube (objectively totally ambiguous) should be seen as left oriented. Similarly, if the average ambiguity for both versions were equal, the A.L. should coincide with the balanced version, and this cube should be judged equally often in each orientation. Experiment I tested the first of these predictions.

For this experiment the 30 A.L. trials were programmed as follows: Trials 1-10 were all between moderately and slightly unambiguous. On Trials 11-20 the trials for one orientation were moderately unambiguous while those for the other were only slightly unambiguous. Finally, on trials 21-30, the trials for the favored orientation were almost totally unambiguous, while those for the unfavored orientation were almost completely ambiguous. For half the subjects the right orientation was favored (less ambiguous); for the other half, the opposite held true. Half the trials in each set of ten were left orientations and half were right.

The 30 A.L. trials were followed by three presentations of the totally ambiguous cube. Between Test Trials 2 and 3 a moderately unambiguous version of the favored side was presented in order to break any set to expect all unfavored versions.

2. Results. The results of this experiment are in line with expectation. Of the total of 30 presentations of the balanced cube (10 subjects x 3 trials), 28 responses were for the orientation opposite to that identified with the A.L. For the subjects with the A.L. established on the right side, the ambiguous version was judged on almost every trial to be in the left orientation; for the judges subjected to operations intended to establish the A.L. in the right perspective, the opposite was the case. In the interest of statistical confirmation, the hypothesis was judged to have held in the individual subject, if he gave two or more responses to the test version in line with prediction. All 10 subjects met this criterion. ($X^2 = 10.0$, df = 1, $P < .01$).

D. Experiment II

1. Procedure. Experiment II was concerned with the effect of a difference in relative frequency of presentation of alternate orientations of equal ambiguity upon the perceived orientation of the balanced cube. Following the logic of Experiment I, it was predicted that the orientation A.L. would be established within the perspective presented most frequently. Thus, for example, if the A.L. was identified with the right perspective, the balanced cube should be judged as facing left.

For Experiment II, the following regimen was used to establish the A.L. On trials 1-14, half the trials were left-oriented and half were right-oriented. On trials 15-30 six were in one perspective and 10 in the other. For half the subjects the favored (the most frequent) perspective was toward the left; for the remaining half toward the right. Over the 30 A.L. trials, the figures were rendered systematically more ambiguous. On the first third they were moderately unambiguous; on the second third between moderately and slightly unambiguous; and on the final third, only slightly unambiguous.

2. Results. Here again, as in Experiment I, expectation is confirmed. Of the total of 30 test trials with the objectively ambiguous figure, 25 of the responses were for the less frequently presented perspective. Nine of the ten subjects met the confirmation criterion ($X^2 = 6.4$, df = 1, $P < .01$).

E. Experiment III

1. Procedure. Experiment I was performed to demonstrate the relation of impressiveness of stimuli over a series of trials upon the judgments of the test figure. In this experiment impressiveness was related to the physical definition of the stimulus. But there are other factors that influence impressiveness. The broad spectrum of operations that serve to emphasize the significant characteristic or characteristics of a stimulus may be assumed to influence the weight of its contribution to the judgmental norm. Reinforcement is one such operation. Experiment III was performed to explore the relationship of reinforcement to this function.

On ten of the fifteen trials on which one of the two perspectives was presented, the subject's response was followed by a confirming comment such as "that's right", "good", or "uh-huh". For half the subjects, the right presentation was reinforced; for the other half, the left. The Experiment II schedule for progressively increasing ambiguity through the 30 A.L. trials was also used in this experiment.

2. Results. The rationale of this experiment predicts that if the right perspective was reinforced during the pre-test trials, the balanced cube would be seen oriented toward the left on the test trials. Results support this prediction. On 25 of the 30 test trials, the perceived orientation was the opposite of that reinforced by the A.L. trials. Nine of the ten subjects met the confirmation criterion ($X^2 = 6.4$, df = 1, $P < .01$).

F. Discussion

The results of the three experiments described in this paper are taken as evidence that perceived figural orientation reflects the influence of patterns of prior input. They suggest that a particular orientation acquires reference or normative status as a function of its average phenomenal clarity, or the relative frequency of its occurrence. Furthermore, its phenomenal clarity would appear to be capable of enhancement by reinforcement.

Certain observations, however, are in order. It would seem, first of all, that the present results contradict findings already in the literature. Several examples bear citing. There are a variety of transactionist demonstrations like those involving the distorted room (6) which indicate that stimulus configurations are seen consistent with past experience. However, it must be pointed out that distorted rooms are not ambiguous stimuli. When viewed under proper conditions they produce retinal configurations which are identical with those produced by normal rooms under the usual viewing condition. They do not differ significantly from the <u>normative</u> stimulus configurations.

Leeper's study (7) with the mother-in-law figure may also be cited. In this study separate groups of subjects were presented one or the other of two alternative unambiguous versions of the test figure prior to viewing the ambiguous version. When this was done, the ambiguous test figure was perceived to be the unambiguous version previously shown. The Leeper experiment differs from those of the present report in one important respect. Its subjects were not permitted experience with alternative figures and, it may be presumed, had no knowledge that alternative versions existed. The adaptation-level model implies the establishment of judgmental norms from <u>varied</u> stimulus input. In the Leeper study experience had no variation at all. And if a judgmental norm was established, it was established in an absolute sense with no range of inputs to provide contrast with the norm. Under such circumstances, one could not expect the test figure to evoke a response other than that of prior training.

The results of Solley and Santos (9) are less easily reconciled with the present findings. These experimenters report that when two completely unambiguous versions of the Necker Cube are differentially reinforced during a series of presentations, the subject more readily and frequently reports the orientation that is reinforced the greater number of times. Again, this study differs in what may be a critical fashion from the present experiments. First, the test trials were interspersed unannounced within the series of unambiguous training trials. This means that testing was carried out at different stages of norm-evolution -- if an orientational norm was generated. Under these circumstances, it might be expected that, at least for the early trials, the norm was unstable or non-existent, with the result that response was required in absolute terms. Now if the occasional test trial in fact was ambiguous, the subject's best response strategy would have been to fall back upon the verbal label used with the greatest success. Furthermore, since, with the exception of the occasional test trial, figures were unambiguously left or right in orientation, the situation may have been structured for the judge as purely dichotomous. Under these conditions, it is not expected that an orientational A.L. would be formed. It is assumed, in contrast, that the operations of the present experiments led to a "dimensionality" of the variable of ambiguity of perceived orientation. It is further assumed that the impression of grades of ambiguity yields an impression of continuous difference in spatial position between left and right orientations. We have therefore conceived of the present results in terms of the A.L. model. Meanwhile, it may be held that they are a matter of simple qualitative contrast. If the standard is identified with one orientation then anything that differs from it, including the ambiguous figure, could be judged in the opposite orientation.

While the present paper was in preparation, Santos and his colleagues (8) reported results which are in line with those of our Experiment II. In two experiments the subjects displayed a strong tendency to perceive the balanced cube in the orientation opposite of that most frequently presented during the training trials. This effect was more pronounced on long-exposure test trials than on short-exposure trials.

A final point concerns a general implication of the present results. Stevens (10) has recently interpreted adaptation-level phenomena in terms of the readjustment of verbal labels used to accomodate particular stimuli within the stimulus series. The shifts in judgment associated with anchor effects, for example, are said to be due to the need to be able to express relative magnitude in terms of the stimulus scale when the range of intensities is increased by the inclusion of the anchor. The present data do not support Stevens' view. It is impossible to deal with complementary orientations in terms of shifting labels in the same way that one can rationalize shifts on an intensive scale.

G. Summary

Three experiments, involving a total of thirty subjects, dealt with perceived orientation of the Necker Cube as a function of patterns of stimulus input prior to presentation of the balanced or ambiguous cube. During a series of pretest trials, the ambiguity of the figure in one orientation, the relative frequency with which it was presented, or the relative frequency with which the orientation was reinforced was varied. It was thus anticipated that this orientation would be established as a standard for further judgments. As predicted, a predominant number of presentations of the maximally ambiguous or balanced cube, were reported to be in the opposite orientation. A reconciliation of these data with previous findings is attempted.

H. References

1. Bevan, W. and Pritchard, Joan F. The anchor effect and the problem of relevance in the judgment of shape. J. gen. Psychol., in press.

2. Brown, D. R. Stimulus-similarity and the anchoring of subjective scales. Amer. J. Psychol., 1953, 66, 199-214.

3. Helson, H. Adaptation-level as frame of reference for prediction of psychophysical data. Amer. J. Psychol., 1947, 60, 1-29.

4. Helson, H. Adaptation level theory. In Koch, S. (Ed.) Psychology: A study of a science, V. I. 565-621.

5. Johnson, D. M. Generalization of a scale of values by the averaging of practice effects. J. exp. Psychol., 1944, 34, 425-436.

6. Kilpatrick, F. P. Explorations in transactional psychology. New York: New York Univ. Press, 1961, Ch. 8.

7. Leeper, R. A study of a neglected portion of the field of learning -- the development of sensory organization. J. genet. Psychol., 1935, 46, 41-75.

8. Santos, J. F., Farrow, B. J., and Solley, C.M. Exposure frequency and the perception of a reversible perspective. Percept. mot. Skills, 1962, 10, 199-209.

9. Solley, C.M. and Santos, J.F. Perceptual learning with partial verbal reinforcement. Percept. mot. Skills, 1958, 8, 183-193.

10. Stevens, S. S. Adaptation-level vs. the relativity of judgment. Amer. J. Psychol., 1958, 71, 633-646.

11. Turner, E. D. and Bevan, W. The simultaneous induction of multiple anchor effects in the judgment of form. Unpublished manuscript.

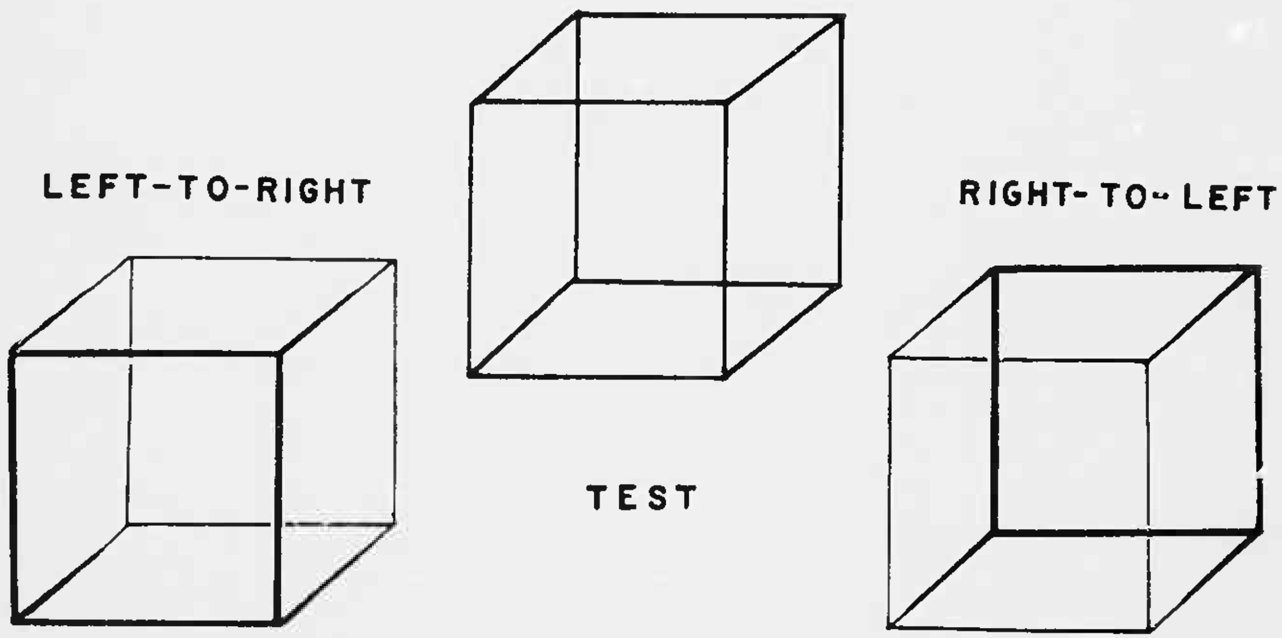

LEFT-TO-RIGHT
TEST
RIGHT-TO-LEFT

Figure 1. The stimulus figures. The figure on the left side is the left-oriented figure; that on the right, the right-oriented. The center figure is the balanced (ambiguous) cube. Not shown is the soft shading on the "near surface" (that bounded by the heavy contours) used to enhance the impression of left or right orientation.

Technical Report No. 6, 1962

The Effect of Subliminal Tones upon the Judgment of Loudness*[1]

William Bevan and Joan Faye Pritchard
Kansas State University

* This report describes work under contract Nonr-3634(01) between Kansas State University and the Physiological Psychology Branch, Office of Naval Research. It is part of a project entitled, "The evolution of perceptual frames of reference."

This paper has been accepted for publication in the Journal of Experimental Psychology.

November, 1962

In a recent paper, Black and Bevan (1960) report that subliminal stimulation, introduced without the subject's knowledge, may effect an increase in the judged intensity of supraliminal series stimuli. Their procedure involved presenting an experimental group electric shocks below the range of threshold sensitivity and interpolated at the midpoint of the temporal interval between successive mild shocks used as series stimuli. Comparison of the responses of this group with those of a control group not presented the subliminal shocks revealed two differences worthy of note: First, the series shocks were judged consistently more intense by the experimental Ss. Second, the slope of the psychophysical function for the experimental group was flatter than that for the control, the elevation of the control judgments being greater for the weaker members of the series than for the stronger. These results suggest that subthreshold stimulation may influence the apparent magnitude of psychophysical stimuli above threshold and that this influence resembles that produced by an anchor above threshold but below the series.

The Bevan and Black data have been confirmed by Goldstone and his colleagues (1962), although the latter found it necessary to modify procedure so that each interpolated subliminal stimulus was temporally closer to the series member it preceded.[2]

The purpose of the present series of experiments was to further establish the Black-Bevan effect as a bona fide anchor effect by demonstrating it for a second sensory dimension. Since electric shock is a noxious stimulus with broad biological significance for the subject, it is conceivable that the subliminal anchor effect may be peculiar to this and similar dimensions. Therefore, the dimension chosen for the experiments herein described was loudness.

Experiment I.

At the time that the Black-Bevan experiment was nearing completion, a pilot experiment involving the judgment of loudness was carried out. This consisted of a counterbalanced design in which all subjects judged the series members under both control (no-anchor) and experimental (anchor) conditions. For half, the control condition occurred on the initial session, with the experimental condition being presented several days later; for the other half, the order was reversed. Tones to be judged were presented through head phones with ambient white noise of moderate intensity present to mask incidental sounds that might serve as anchors above the series.

The results plotted as 2 functions separated by a fraction of a category-unit's difference at the low end of the series and converging on a common value at the upper end of the series. However, the relative position of the 2 curves was opposite to expectation; the judgments in the presence of the subliminal stimulus were less intense than those of the control condition. While the between-conditions source of variance was not significant, the difference in slopes (conditions x stimuli interaction) was highly reliable ($P < .001$).

Since a preliminary experiment with shock had shown already that order of presentation influences anchor effectiveness, it was conjectured that these results might reflect the operation of this variable. Accordingly, Experiment I is a repetition of the loudness experiment using the simple two independent-groups design.

Subjects. The Ss were 30 undergraduate students enrolled in Introductory Psychology: 7 men and 8 women made up the control group while 8 men and 7 women constituted the experimental group. None had previous experience in psychophysical experiments.

Apparatus. The stimulus tones were generated by a Hewlett-Packard Audio-oscillator, Model 200CD and presented binaurally to S through Telephonics Type TDH-39 high fidelity earphones. The intensity of the stimuli (1000 cps tones) was set manually on each trial by the experimenter by means of a Hewlett-Packard Model 350 B Attenuator. The duration of each presentation as well as the intervals between presentations was controlled by 2 Hunter Model 111 C Golden Silence interval timers appropriately connected for recycling. The elimination of clicks associated with onset and offset of the tones, as well as control of their rise and decay times, was accomplished through use of a Grason-Stadler electronic switch. The output of a General Radio Type 1890A Random Noise Generator, amplified through a Bell Model 2122-C amplifier and presented through an Oxford 8-in. speaker provided the ambient sound screen.

Procedure. Each S was tested individually. The first step in the experimental procedure involved determining his loudness threshold in the presence of a continuous noise level of 71 db. A variation of the method of limits was used, the threshold being taken as the median of 15 momentary estimates obtained with an ascending series alone.

S was next instructed in the judgment of sensory stimuli using the rating-scale version of the absolute method. He was given 9 categories, varying from very, very soft through medium to very, very loud, but was told to use as many categories as possible and to add categories when appropriate. Finally, it was explained that loud and soft had reference only to the stimuli he would be presented during the test session.

Each S received 20 presentations of each of 5 series intensities, with the order of presentation random for 5 successive blocks of 20 trials each. The physical intensities used were set with reference to the individual S's absolute threshold. The weakest member of the series was 5 db above this value, with the additional stimuli represented by successive increments of 5 db. The subliminal anchor was set at 5 db below threshold and was interpolated without S's knowledge. Care was taken to eliminate any member from the experimental group who assigned a category value on one or more presentations of the subliminal stimulus or who either voluntarily or upon questioning at the end of the test session gave evidence of being aware of its presence. The series-stimuli were presented at intervals of 10 sec. for a duration of 1 sec. The subliminal stimulus was introduced at the midpoint of the presentation interval, i.e., 5 seconds after the onset of each series-stimulus. Medians of the category judgments made for the several series-stimuli by the individual Ss served as data for statistical analysis. Psychophysical functions for the 2 groups were constructed from the means of these medians.

Results. The results of Experiment I are seen in Figure 1. It may be noted that there are 2 non-overlapping functions with maximum separation at the low end of the series. Statistical evaluation by means of a trend test indicates the separation of the 2 curves to be reliable ($F_{\text{between groups}}$ = 52.22, df = 1/84, $P < .001$) while it fails to confirm the asymmetrical anchor effect ($F_{\text{between slopes}}$ = 1.32, df = 1/84, $P > .05$). These data are in line with those of the pilot experiment. The judgments of the experimental group, contrary to expectation, are less intense than those of the control, but, since the design involved independent groups, cannot be attributed to an interaction involving order-effects. Two other possibilities suggest themselves. It is conceivable that the masking noise, instead of screening out incidental anchors, itself constitutes an anchor above the series, interacting with the experimental conditions to produce the obtained results. On the other hand, the subliminal anchor, being relatively impotent and close to the series, could have effected <u>assimilation</u> rather than the more frequently-observed <u>contrast</u> anchor-effect. Therefore, Experiment II was undertaken to check these possibilities.

Experiment II

Experiment II was essentially a repetition of Experiment I under circumstances which allowed the omission of the ambient white noise. This was accomplished by conducting the test session in a specially-constructed test chamber, capable of attenuating extraneous noises by approximately 60 db., with the subject wearing headphones mounted in ear muffs capable of the further dampening of outside sound by 45-60 db.

<u>Subjects</u>. Eighty Introductory Psychology students, 20 men and 20 women in the control and experimental groups respectively, participated as <u>S</u>s.

<u>Apparatus</u>. The apparatus was that of Experiment I except that the Telephonics Type TDH-39 phones were replaced by Willson Sound Barrier ear muffs containing Stromberg Carlson high-fidelity phones.

<u>Procedure</u>. The procedure of Experiment II was identical with that of Experiment I except that the physical intensities of all stimuli were the same for all subjects. The series intensities were 26, 31, 36, 41 and 46 db S.P.L. respectively. The subliminal stimulus was 16 db S.P.L. Each <u>S</u> in the experimental group was carefully observed while being tested and was interrogated at the end of the session. Twenty-five evidenced knowledge of the presence of the subliminal input and were disqualified as <u>S</u>s. These are in excess of the 80 identified above as <u>S</u>s.

<u>Results</u>. The elimination of the sound screen and the insulation of the subject from extraneous noise yielded data clearly different from those of the pilot study in Experiment I (cf. Figure 2). It thus seems reasonable to attribute the reversal in the relative positions of the experimental and control curves obtained in these earlier investigations to the anchor-like intrusion of the masking noises. Meanwhile, the results are not definitive confirmation of the proposition that subliminal stimulation may behave like an anchor below the series. Although the E-group values are greater than corresponding C-group values for the 3 loudest members of the series, the curves seen in Figure 2 intersect at a value approximately equal to that of stimulus 2 and the reversal for the weakest series member suggests a slight assimilation effect. Analysis of variance of the data of Experiment II indicates no reliable between-groups difference ($F_{\text{between groups}} = .33$, df = 1/76, $P > .05$), but a significant difference in slopes ($F_{\text{group x stimuli}} = 4.75$, df = 4/304, $P < .01$).

Experiment III

The purpose of Experiment III was two-fold: to seek a more effective set of stimulus conditions for the production of the subliminal anchor effect, and, as part of this, to explore the result of changing the temporal position of the anchor within the presentation interval. Goldstone, Goldfarb, Strong and Russell (1962) indicate that when the subliminal anchor effect for shock could not be obtained with interpolation at the midpoint of the presentation-interval, it was possible to evoke it by moving the anchor back so that it preceded each series member by 2 rather than 5 seconds. Similarly, Maruyama (1957), in a study of intermodal relations, interpolated a tone either immediately following the standard visual stimulus or just prior to the comparison stimulus and found that the tone influenced the brightness judgment only when it occurred in the latter temporal position. Accordingly, it was decided to examine the subliminal anchor effect with the anchor at several positions within the presentation interval. Since inspection of Figure 2 suggests a weak and inconsistent anchor effect, it was assumed that this effect might be enhanced by increasing the scalar distance between series and anchor magnitudes.

Subjects. Three groups of 20 subjects each, 10 men and 10 women, represented the experimental condition. The control group of Experiment II was used to provide control data for this experiment. As in the case of Experiment II, subjects indicating cognizance of the subliminal inputs were eliminated and are in excess of the 60 designated as subjects.

Procedure. The subliminal stimulus was introduced early in, at the middle of, or late in the presentation-interval. For the early group, its onset followed that of the preceding stimulus by 3 seconds, for the middle group, 5 seconds, and for the late group, 7 seconds. Two additional Hunter timers were added to make possible the temporal programming. The subliminal stimulus was set at 11 db S.P.L.

Results. The results of Experiment III are seen in Figure 3. It will be noted that in every case the curve for the experimental group lies above that of the control. In every case the difference between conditions is statistically reliable ($F_3 = 5.09$, df = 1/76, $P < .05$; $F_5 = 7.97$, df = 1/76, $P < .01$; $F_7 = 4.44$, df = 1/76, $P < .05$). Thus, it would appear that an appropriate set of experimental conditions had obtained and the subliminal anchor effect observed by Black and Bevan for shock was confirmed for loudness. Similar support for the generality of the subliminal anchor effect is provided by a recent study by Boardman and Goldstone on the judgment of size. Using a visual-masking technique with tachistoscopically-presented discs, these experimenters supplemented the psychophysical series with an anchor disc either larger than or smaller than the series members. Although the differences were small, the judgments for the group receiving the small anchor were consistently greater for the experimental than for the control condition; while those for the group receiving the large disc were, as expected, less than the control judgments.

Figure 4 is a graph of the adaptation-levels (the stimulus magnitudes corresponding to the judgment of medium) for the experimental groups of Experiment III expressed in db. It is clear from the zero slope that the subliminal anchor effect did not differ over the range of interpolation-times used. This, of course, does not mean that interpolation time is not an important general consideration, but only that for the particular experimental conditions employed in Experiment III, it failed to produce a differential effect.

Discussion

The aim of the experiments reported in this paper was to explore the generality of the subliminal anchor effect reported by Black and Bevan (1960) for the sensory dimension of electric shock. A review of the results, taken as a whole, leads us to conclude that the effect is a genuine one. Beyond this, there are several general comments that warrant mention.

The experiments on subliminal anchors are part of a larger program of studies dealing with the problem of relevance in sensory judgment. More specifically, this program is directed toward identifying the properties that identify stimuli that are pooled by the judge and therefore determine the scalar value of an adaptation level in contrast to those which exert no influence upon the judgmental norm. The present results are consistent with the Black and Bevan conclusion that the absolute threshold need not be a limiting condition in pooling.

Two characteristics of the Black and Bevan results support the conclusion that the subliminal anchor effect is a genuine anchor effect: (a) the judgments of series stimuli with the anchor interpolated were significantly greater than when it was absent; and (b) the subliminal anchor effect is greatest for stimuli which are at the low end of the series and thus also nearest the position of the anchor in the stimulus scale. Comparison of Experiments II and III adds a third characteristic. (c) The subliminal effect was enhanced by changing the intensity of the subliminal stimulus. This involved _increasing_ the scalar distance between the anchor-designate and the series members while at the same time _reducing_ its physical magnitude and taking it _farther_ below threshold.

Stevens (1958), in his discussion of the concept of adaptation level, has suggested that when one goes beyond the problem of color constancy and contrast, A-L phenomena may be attributed to the judges' shift of category names to provide for the relative position of the series stimuli in an extended range provided by the introduction of the anchor. The subliminal anchor data, while they do not invalidate this explanation for certain anchor phenomena, raise questions about its generality.[3]

SUMMARY

A pilot study and three experiments, involving a total of 144 subjects, were conducted to determine if a subliminal anchor-effect, already reported for the sensory dimension associated with mild electric shock, could be demonstrated also for the dimension of loudness. After some manipulation of the experimental conditions, it was found that the introduction of a tone below audible threshold was accompanied by an elevation in the judged magnitude of series stimuli above threshold. It is noteworthy that the anchor effect was finally produced by increasing the scalar distance between the series and the anchor-designate, although this involved decreasing the physical intensity of the anchor and taking it farther below threshold. Implications of the results for adaption-level theory are suggested.

REFERENCES

Black, R. W. and Bevan, W. The effect of subliminal shock upon the judged intensity of weak shock. Amer. J. Psychol., 1960, 73, 262-267.

Boardman, W. K. and Goldstone, S. Effects of subliminal anchors upon judgments of size. Percept. mot. Skills, 1962, 14, 475-482.

Goldstone, G., Goldfarb, Joyce, Strong, J. and Russell, J. Replication: The effect of subliminal shock upon the judged intensity of weak shock. Percept. mot. Skills, 1962, 14, 222.

Maruyama, K. The effect of tone on the successive comparison of brightness. Tohoku psychol. Folia, 1957, 15, 55-70.

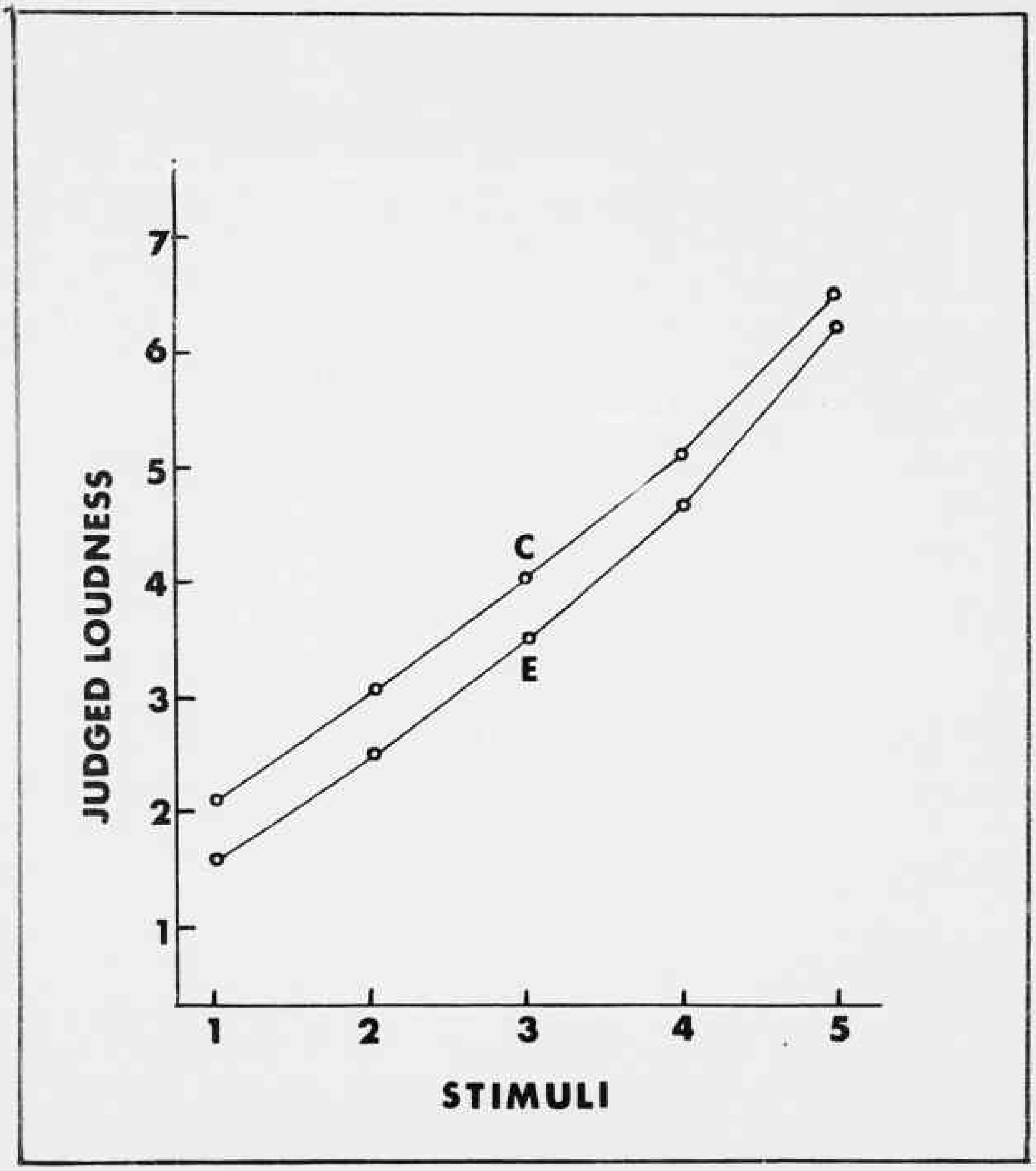
Stevens, S. S. Adaptation-level vs. the relativity of judgment. Amer. J. Psychol., 1958, 71, 633-646.

FOOTNOTES

1. The authors are indebted to Miss Josephine Baker and Mr. Lonnie D. Whitehead, who performed the pilot experiment as part of an advanced course in experimental psychology. They are also grateful to Messrs. Gherry Harding and Jerome Walker, who jointly served as experimenters for Experiment I while participants in an N.S.F.-sponsored Undergraduate Research Participation Program during the summer of 1960, and Mr. James Haines, who helped with data collection for Experiment II. The junior author, J. F. P., served as experimenter for Experiments II and III. Experiments II and III were performed under Contract Nonr 3624(01) between Kansas State University and the Physiological Psychology Branch, Office of Naval Research.

2. It is appropriate to note that these subliminal anchor effects are relatively subtle and require careful manipulation of the experimental conditions in order that they be found in evidence. Consideration must be given to both the temporal and intensive proximity of the anchor to the series members, its duration, and the relative frequency of its occurrence. Since anchor effectiveness is easily masked by other effects, the present investigators have found that the simple independent groups design, with groups of at least 20 subjects each, is most efficient.

3. While the present paper may not be an appropriate place for an extended discussion of the general implications of the subliminal anchor effect, one methodological issue bears suggestion. It is widely held that the strongest form of scale, the ratio scale, has been achieved for certain sensory dimensions through the use of such psychophysical methods as that of fractionation and magnitude estimation. These methods are based on the assumption that the threshold constitutes the absolute zero of the scale. The subliminal data, since they indicate that magnitudes assigned to supra-threshold stimuli may vary in the presence of subliminal inputs, would appear to cast doubt on the possibility that true ratio scales have been constructed. This is a matter that warrants careful examination.



JUDGED LOUDNESS
1
2
3
4
5
6
7
C
E
1
2
3
4
5
STIMULI

Fig. 1. Judged loudness of tones for experimental and control groups of Experiment I.

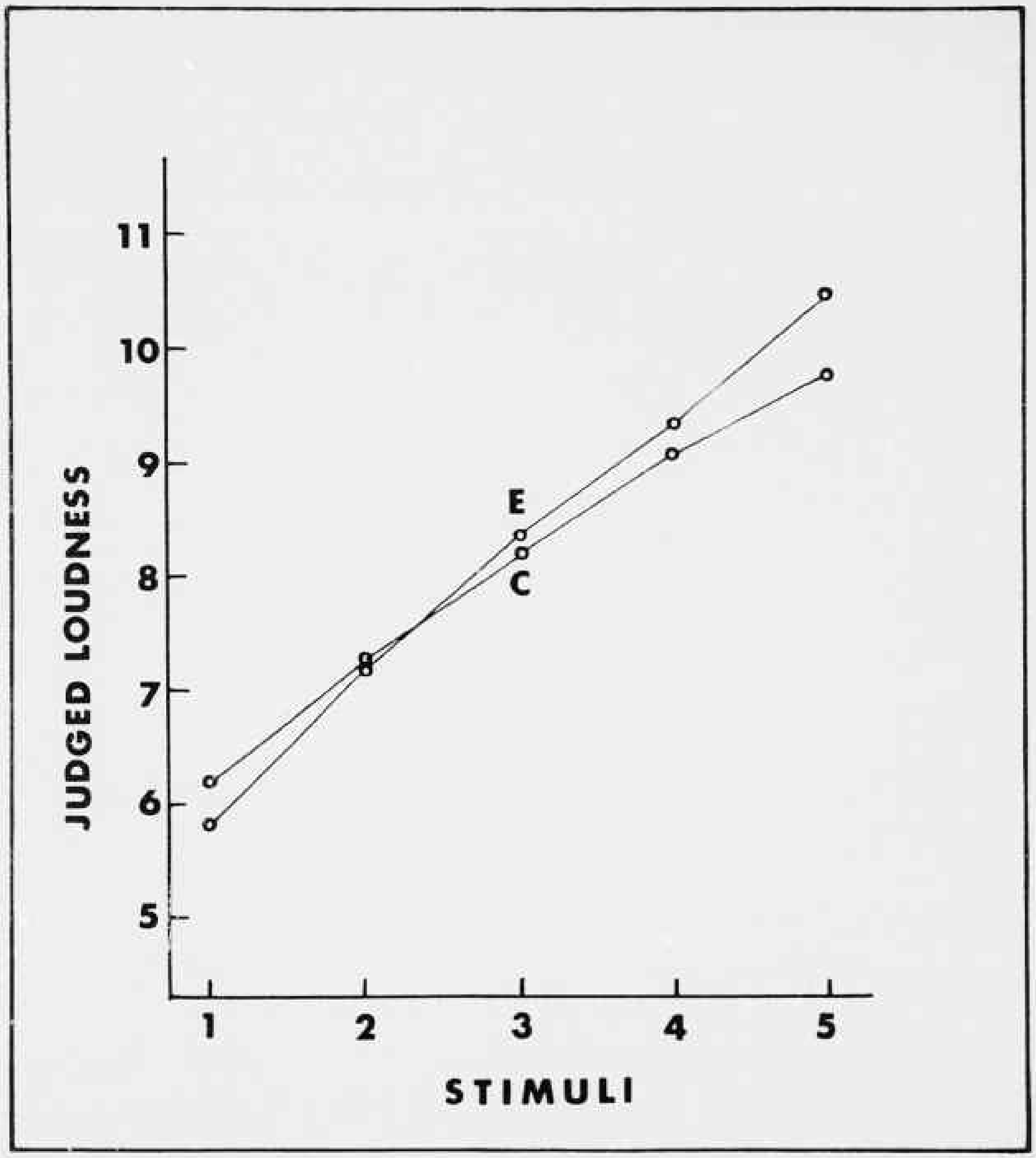
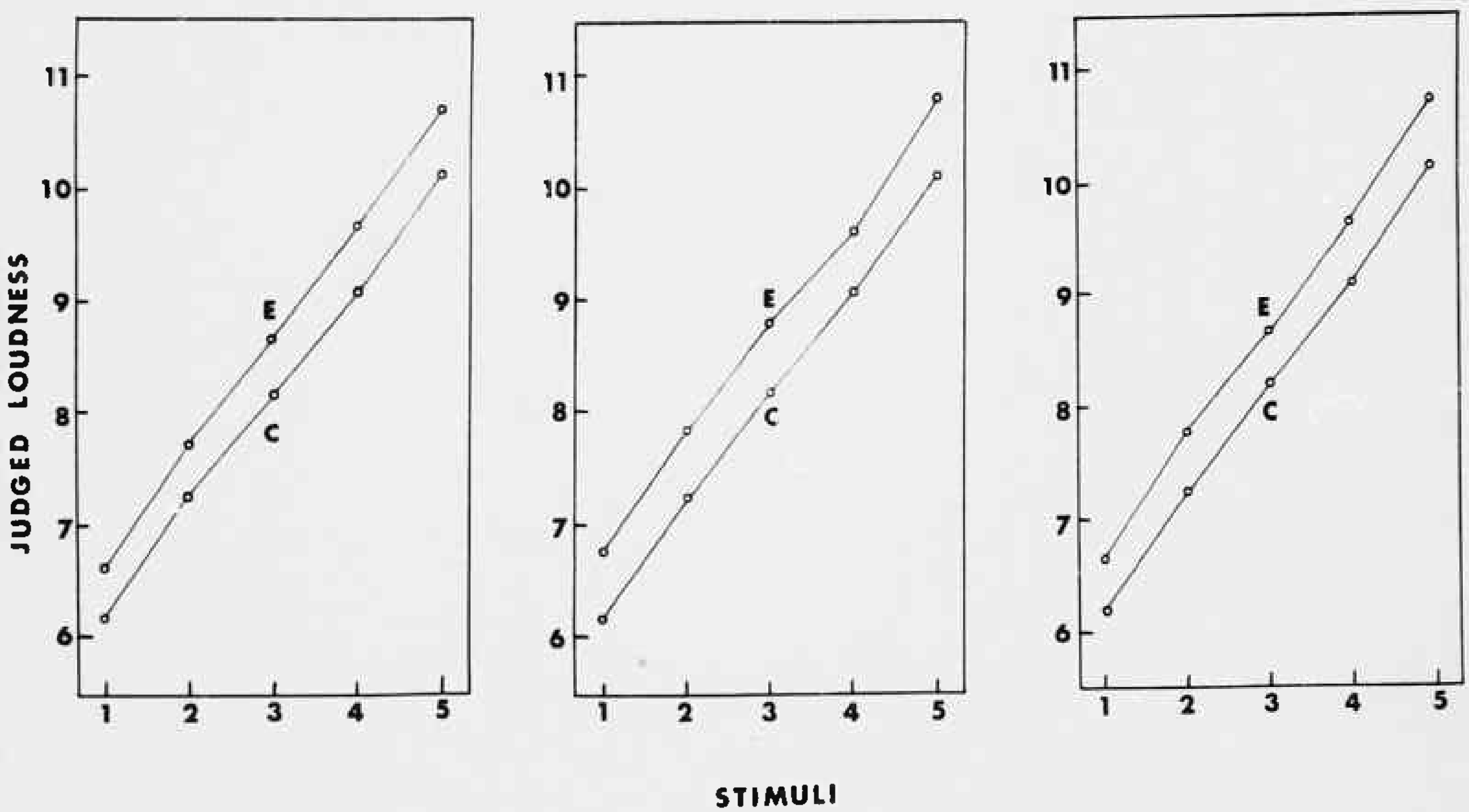
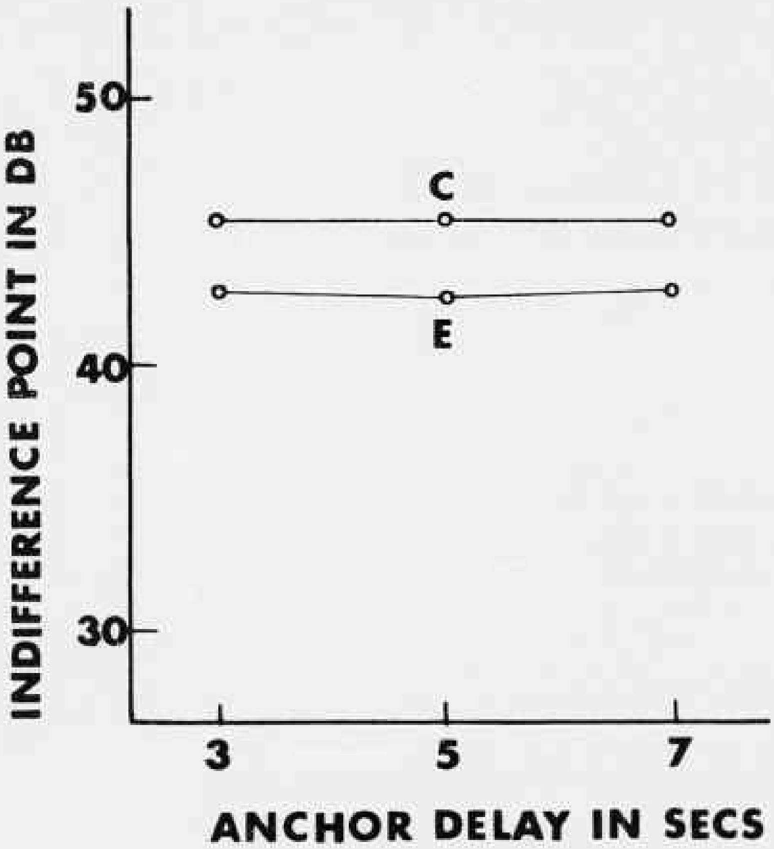

11
10
9
8
7
6
5
JUDGED LOUDNESS
E
C
1
2
3
4
5
STIMULI

Fig. 2. Judged loudness of tones for the control and experimental groups of Experiment II.



JUDGED LOUDNESS
11
10
9
8
7
6
E
C
1
2
3
4
5
STIMULI

Fig. 3. Judged loudness for the groups of Experiment III. The left-hand graph presents the data obtained when the anchor occurred early (3 seconds after onset of the preceding stimulus) in the presentation interval; the center graph, the data when the anchor occurred at the midpoint of the interval (5 seconds after onset of the preceding stimulus); and the right-hand graph the data when the anchor occurred late in the presentation interval (7 seconds after onset).



INDIFFERENCE POINT IN DB
50
40
30
C
E
3
5
7
ANCHOR DELAY IN SECS

Fig. 4. Magnitude of the anchor effect as a function of the position of the anchor in the presentation interval. The medium loudness judgment is represented by the ordinate; the temporal proximity of the anchor to the preceding series stimulus on the abscissa.

Technical Report No. 7, 1962

On the Approach of the Experimental Psychologists[1]

William Bevan

Kansas State University

1 This essay was prepared under contract Nonr-3634(01) between Kansas State University and the Physiological Psychology Branch, Office of Naval Research. It is related to a project entitled, "The evolution of perceptual frames of reference."

It will appear at a later date in "Psychology and the Philosophy of Science", B. B. Wolman and E. Nagel (Eds.), Basic Books, Inc.

November, 1962

Introduction

There are a number of ways in which the several activities collectively identified as performing experiments may be figuratively characterized. One is a game in which the goal of the experimenter-player is to confirm anticipated relationships in as expeditious, economical, and aesthetically-satisfying a fashion as possible. The basic rules are rather simple and easy to learn; the techniques vary in complexity depending upon the phenomena under consideration. The intuitive strategies that separate the highly original, creative investigation from others come only after deep immersion in a problem area, but otherwise defy understanding. Two criteria are predominant: all solutions must be empirically-grounded and all operations, both procedural and conceptual, must be logically coherent. It is usually expected that these two requirements will complement each other. However, when they come into conflict, the experimenter, when he has satisfied himself that his procedures are appropriate, is committed to granting precedence to the empirical criterion and denying certain of the assumptions which gave rise to his investigation.

This chapter is an attempt on the part of a work-a-day experimental psychologist to describe the psychology of doing experimental studies of behavior. It is not a treatise on the philosophy of science, although much that is dealt with comes under the scrutiny of this separate discipline. My concern with certain philosophical issues is motivated by practical considerations related to getting a research job done. Issues of a purely systematic nature, often subtle beyond my comprehension, I leave to the philosophers. Neither is this paper a set of instructions on how to perform experiments. This task is left to the writers of laboratory manuals and the design statisticians. Rather, it is my purpose to identify the assumptions and attitudes that I find present to guide my own work and which I believe I hold in common with other experimental psychologists. The chapter consists of four major divisions: a delineation of the outlook and assumptions shared by behavior scientists, a description of their materials, some underlying properties of method, and a discussion of the art of doing experiments in psychology.

The Point of View of the Experimental Psychologist

Each experimenter takes certain things for granted when he undertakes to perform experiments: Some of his assumptions are explicit, others are implicit. They deal with the conceptual orientation (The psychologist calls this his theory, although, in terms of the criteria provided for theory-making by the systematists it may be poorly constructed theory, indeed.) he finds useful, the general nature of the phenomena he studies (These are superordinate to any particular assumptions dictated by his theory and thus would be held in common by experimental psychologists regardless of theoretical persuasion.), and the nature of data and the properties of his method. Assumptions of the first sort will more than likely be explicitely stated. Many relating to the latter two categories go unrecognized, and one of the most important exercises any investigator, behavioral or otherwise, may engage in involves teasing out for scrutiny those that lie hidden in his thinking.

This chapter does not concern itself with assumptions relating to the content of specific theories. Rather, it deals with the experimental psychologist's broader biases concerning the material he wishes to study and the most fruitful way of going about this activity. The following three propositions represent the nature of his bias concerning his subject matter.

1. Behavioral phenomena are physical phenomena. Almost without exception present-day experimental psychologists agree to the correctness of this statement. That they all agree on what is meant by it is another matter.

One very influential interpretation of physical phenomena is identified with early Behaviorism. From Watson to the present many psychologists have insisted that behavioral events are in principle definable in terms of the substantive concepts that characterize the physical sciences. To such persons, for example, conditioning is thought of as electrochemical change at the synapse (Johnson, 1927), memory as involving the elaboration of molecular lattices in the brain tissue (Katz and Halstead, 1950), and emotion as the disruption of cortical organization (Hebb, 1949). Though concern for physiological mechanisms almost disappeared from the intellectual currency of psychologists in the thirties, forties and early fifties -- due primarily to the vigor and impressiveness of such persons as Tolman, Spence, and Skinner -- the return, in the 1950's, to an interest in physiological psychology and the explosive increase of this interest is testimony to the depth of this conviction.

At the same time, there has been a persistent uneasiness about this interpretation. Originally, it stemmed from the problem of the status of introspective observations. Watson had equated the observable with the physiological and had declared that only the observable was a proper subject of scientific inquiry. The existence of subjective events was even denied. While there has been much parochial concern among the methodologists of behavioral science concerning the status of introspective data, the concern of the working experimenter has been motivated by a more practical consideration: accepting the criterion of intersubjectivity at face value makes certain important areas of interest inadmissible to scientific inquiry. Allport (1947) summed up this problem in his charge that psychology's addiction to the machine model, and its attendant preoccupation with lower animals as ideal subjects, has resulted in an inability on the part of systematic psychology to adequately accommodate the most important aspects of human behavior -- moral nature and social skills. This sentiment has been expressed more recently by Koch (1961), who insists that major psychological problems require the degree of sensitivity to the subtleties of individual experience identified with the humanities. The physicalistic orientation of psychologists also has been accompanied by a preoccupation on their part with situational variables and an ignoring of value properties intrinsic in behavior itself. Hebb (1952) has admitted that present-day psychological theorizing is inadequate, but he argues that this does not result from the use of the physiological idiom, but rather from the use of an already-dated set of physiological concepts. In his view (1948), the central problem of psychology is the problem of thought, and the peripheralistically-oriented physiology of the 1920's prevalent among many present-day psychologists is not conceptually adequate to handle the matter of behavioral processes identified with higher brain centers.

The central concept identified with the methodological approach of behavioral science is that of intersubjectivity, i.e., the requirement that all observations be capable of confirmation by more than one observer. This has meant to most psychologists of the past several decades that their proper subject matter is circumscribed by the operations that yield dispositional concepts and those that constitute empirical-logical links between the hypothetical and the dispositional levels. That is to say, one may deal only with events of a public character or processes directly linked to them. Zener (1958) meanwhile has suggested that the criterion be broadened from one of intersubjective agreement to one of repeatibility of obtained correlations between sets of conditions and experiential report. This is not far from what logical positivists themselves are now saying. Carnap (1956, p. 70-71), for example, comments: "Although many of the alleged results of introspection were indeed questionable, a person's awareness of his own state of imagining, feeling, etc., must be recognized as a kind of observation in principle not differing from external observation and therefore as a legitimate source of knowledge, though limited by its subjective character." If the admissibility of intrapersonal observation as objective observation is recognized, then the methodological problem becomes one of technical competence, not logical constraint, and the experimentalist in psychology faces the need to enhance each observer's range of sensitivity to his own reactions and his precision and reliability in transforming these into data.

It is apparent that the public-private dichotomy present in the original Behavioristic criterion for defining the admissibility of data results from some confusion between the nature of observations and the nature of data. Observations, whether they be of one's own dreams or involve the pointer readings so frequently mentioned by the operationists, are always, in the final analysis, subjective and private. Data, on the other hand, whether they describe dreams or pointer readings, are always public in nature, for they are human conventions devised for the communication of information derived from observations. The matter, therefore, is not one of public versus private realms of observation. Rather, it is a matter of efficiency -- i.e., the completeness, precision, and reliability -- with which observations are transformed into data.

This all suggests that behavioral phenomena are physical phenomena in quite another sense -- that is, they have the same formal properties that the phenomena of physical science have. They yield data which are objective and replicable and they are accounted for in terms of general principles, which, like physical laws, are assumed to be invariate for all situations to which they apply. Thus, as Adams (1954, p. 66) has put it, "A sentiment or a psychical system or a superego is just as physical a notion as a cell assembly, an engram, a reflex arc, or even an atom."

The problem of generality has a special reference for the psychologist, for he is faced with extrapolating from one phyletic level to another. A study of the interpretative activities of the experimental psychologist suggests in these activities that he takes recourse to analogy. To the extent that he can assume that the actions, past history, and the biological nature of another organism in a particular situation resemble what he concludes his own would be in that situation, he will identify behavior in terms of his own behavior and experience. The confidence that he attributes to these interpretations will vary directly with the degree of judged similarity between himself and the organism observed.

2. Behavioral phenomena are complex. It has become increasingly clear in the last decade that even the simplest behavioral phenomenon represents a complex of determining processes. This recognition is seen in many quarters: e.g., in Hebb's plea (1952) for an enlightened use of physiological models, in his identification (1949) of thought as the central problem of psychology, and in his theoretical interpretation (1949) of the course of perceptual learning; and in Koch's criticism (1961) of extant theories of motivation and his conviction that an understanding of the major problems of psychology require levels of experiential sensitivity identified with the humanities. The current interest in perception also typifies this awareness of complexity. Classical

introspectionism concentrated its attention upon the sensory processes, which it viewed as the simple building blocks of perception. Its studies of these phenomena were limited exclusively to the identification of their physical stimulus correlates. In contrast to this approach, since the late 1940's there have been widespread interests in perception that have involved a consideration of its motivational determinants, and the effects of both specific practice and generalized learning upon it. Other interests in perception have involved early experience, social factors, personality organization, unusual environments (for example, those involving marked enhancement or reduction in level of stimulation as well as imbalances in the atmosphere and extremes of temperature), the influence of damage to the central nervous system and the subtle and diffuse effects identified with drugs and changes in endocrine status. Traditional psychophysical theory, in both its Fechnerian and modern version, assumes that judgmental magnitudes are adequately specified by reference to corresponding physical stimulus magnitudes (Guilford, 1954, Ch. 2; Stevens, 1957). In contrast, Adaptation-Level Theory (e.g., Helson, 1959) formally recognizes three classes of determining variables: (a) the focal stimulus of traditional psychophysics, (b) the immediate background, of major concern to Gestalt psychology, and (c) residual stimulation, which includes a wide range of determinants, both identified and non-identified, but which, it may reasonably be assumed, are predominantly of central origin.

The awareness of complexity on the part of the present-day experimental psychologist is seen in certain broad attitudes and approaches. He has of late shown more reticence about extrapolating experimental results across species and situations. In his animal experiments he has to an increasing degree chosen species closer to man on the evolutionary scale than the white rat, and in his theoretical exercises he no longer views the rat as a simpler version of the human machine. After many years of relative neglect, Brunswick's notions (1956) of representative design and ecological validity have caught hold and more serious attempts to simulate in the laboratory the situation to which the experimenter wishes to generalize his results are in evidence -- most frequently this means "the situations of everyday life". Greater concern is evidenced for the effect upon his data of procedural variables such as order-effects, instructions, individual differences in experimenter technique, personality characteristics of both experimenter and subject, sex differences, and incidental situational factors.

Still, the profound influence of the experimenter upon experimental data is only partially formally recognized (McGuigan, 1962). At the same time, use of complex experimental designs is widespread by all but the most committed of the students of operant conditioning to control incidental variables as well as to identify subtle behavioral determinants. [These latter are more inclined to confine themselves to situational variables and, at least in prototype, eschew the goal of general predictability (Skinner, 1938).] Even when interest is confined to the role of the focal stimulus in behavior, this now tends to be marked by a greater sensitivity to subtle relations in the stimulus array -- as in Gibson's analysis (1950) of the stimulus determinants of impressions of form in depth -- than has heretofore been the case.

However, the recognition of complexity at the empirical level should not be confused with the goal of conceptual simplicity which the experimental psychologist shares with other scientists. It would appear that even the most empirically-minded scientist is not content with the production of data alone. Observations must be fitted together into an explanatory scheme that is logically adequate to the domain into which the data fall and which at the same time is neat, deft in contrast to clumsy, and pleasing. Why the conceptual scheme must be simple and aesthetically satisfying are profound psychological problems that we do not purport to answer here. Suffice it to say that aesthetic satisfaction appears to be the goal of all creative work and this aspect of the problem is extra-logical. At the same time, the task of identifying what shall constitute a conceptually simple explanation involves certain logical considerations. If we examine many of the theories that have been extant in the past several decades one is led readily to the suspicion that the common practice has been to begin with a point of view and then to work to preserve its integrity by the reclassification and exclusion of data that it fails to accommodate and/or by the introduction of auxiliary hypotheses. Popper (1959) has made much of the subtlety of the epistomological problem of simplicity.

How knowledgeable the working scientist must become of the logician's analyses is a matter of practical concern. At present, he regards as most simple the theory that adequately accommodates with the smallest number of assumptions the largest number of phenomena within its domain. Furthermore, the assumptions are expected to be <u>general</u> and <u>a priori</u> rather than <u>special</u> and <u>ad hoe</u>.

3. <u>Behavioral data may be scaled or otherwise subjected to quantification</u>. The physical science ideal and the goal of scaled data are synonymous. Adopting the approach of the physical sciences as a methodological model is tantamount to a committment to quantification and to measurement as the physical sciences achieve it. In its most efficient form this means identifying observations as numbered positions along a single dimension which may be summarized by the operations of arithmetic.

For the experimental psychologist it is not so much a matter of whether or not he can quantify. This is ultimately a matter of the convention he will use to describe his observations and it will be, in some respects, at least, arbitrary once his measurement model is selected. Rather it is a matter of what level of quantification is most efficient with the phenomena he has under study; that is to say, what form of quantification communicates the maximum amount of information without distortion.

What levels are possible with behavioral data can be understood at least partly by reviewing the kinds of operations the experimental psychologist performs when he collects data. The great diversity of data that characterizes behavioral science can be identified in terms of two kinds of quantification: (1) counting and (2) measures of intensity. The rational processes involved in the formalization of data are either discriminational or judgmental, although in the case of any particular datum, which is involved may be unclear. In the case of the discriminational, the experimenter need only identify the occurrence or non-occurrence of an event (the event may itself be the occurrence or non-occurrence of a difference among events). In contrast, the judgmental requires the assessment of degree of magnitude or the identification of scalar position relative to some criterion dimension.

The ultimate form that data take is not perfectly correlated with the nature of the collection process. Some counting data presuppose judgmental evaluation of several individual occurrences before discriminations can be made. On the other hand, some measures of intensity are derived from counting data and certain assumptions about the relationship of the frequency distribution of these data to an underlying intensive continuum. For example, if one is interested in quantifying the preponderance of cigarette smoking in a particular group, he will undoubtedly count the number of smokers involved. But first he must set criteria for the identification of smokers and classify each case before a count can be made. Meanwhile, if one wishes to estimate the maximum height an individual can jump, one has him scale a bar, beginning at a height he can successfully negotiate on 100% of the trials allowed. This procedure is then repeated with the bar a fixed increment higher on successive trials until his 100% success is transformed into 100% failure. The scale value corresponding to this frequency count then becomes a reasonable estimate of the magnitude sought.

The complementarity of counting and intensive measures relates to the fact that, almost without exception, scientific data are representational. This is the case regardless of whether the conventions the scientist uses to describe the data come from his own experience or are acquired from physics. Rate of bar-pressing in the white rat may, depending upon the circumstances, indicate sensory discrimination, strength of drive, or level of learning. The only instance in which behavioral data constitute direct quantification is when the psychologist is interested in performance phenomena per se and here the nomenclature is the same as that of molar physics: latency, duration, speed, magnitude of output, efficiency, etc. Even in the case of performance phenomena, none of the measures involved are fundamental in the same sense that certain physical dimensions like distance, mass, and time are fundamental, i.e., independent of other measures for their specification (Campbell, 1928).

The experimental psychologist was committed to quantification over one hundred years ago by Herbart, by Weber, and by Fechner. In his Psychologie als Wissenschaft, Herbart argued that psychology was a mathematical science. Herbart's mathematics, however, was not measurement mathematics but the elaboration of rational equations in the service of a particular metaphysics, and it has not survived. But, along with Weber's empirical principle linking the difference threshold to stimulus magnitude, it did provide the basis for Fechnerian psychophysics, the prototype of measurement in psychology.

A metaphysician and mystic as well as a physical scientist, Fechner believed that if the psychophysical relationship could be quantified, the mind-body problem could be resolved. The solution he proposed he referred to as Weber's Law. This he deduced from several assumptions: That for every physical magnitude within the sensible range there exists a corresponding sensory magnitude; that the origin of sensory dimensions coincides with the absolute threshold; and that from the Weber ratio it may be properly inferred that the difference limen (or any small increment in the stimulus) corresponds to a constant distance on the sensory dimension. Since differential sensitivity is proportional to stimulus magnitude, the psychophysical relationship is inferred to be a logarithmic one:

$$\text{Judgment} = C\,(\log \text{Judged Stimulus} - \log \text{Absolute Threshold Stimulus}).$$

The construction of a psychophysical function then would involve basically two measures; a determination of the absolute threshold and the jnd, and the cumulation of jnds from the origin over the sensible range.

Fechner denied the possibility that sensations could be measured directly. All one could do was to count the frequency with which sensations (or sensory differences) were present or absent. This is equivalent to constructing a set of conditional probabilities relating a series of stimuli to a criterion response. Thurstone (1927a, 1927b) turned this reasoning around. The true judgmental magnitude corresponding to any stimulus magnitude is held to be the mode of a distribution of responses given to that magnitude. Assuming that the response dispersion for any stimulus is normal, the discrimination of this stimulus from any other reflects the degree of overlap of the two dispersions involved. When overlap is small the two stimuli are easily discriminated; when it is great, they are readily confused. Since the degree of dispersion may be expressed in terms of any common measure of variability, the specification of sense distance becomes possible. This is formally stated as the law of comparative judgment:

$$\text{Judgment}_1 - \text{Judgment}_2 = Z_{12}\sqrt{\sigma^2_1 + \sigma^2_2 - 2r\sigma_1\sigma_2}$$

where Z is the proportion of times one stimulus is judged higher than the other, and σ^2_1 and σ^2_2 are the variances of the perceived values produced by the two stimuli.

The solution of this equation thus replaces the jnd as Thurstone's measure of distance. Its presumed advantage is that it allows for the construction of psychological continuua regardless of the nature of the stimulus scale -- indeed, even in cases (e.g., attitudes, beliefs, etc.) where a physical continuum evades ready identification or logically does not exist.

Of late, both Fechner's and Thurstone's approaches to scaling have come in for serious criticism. For some time the Weber fraction has been known not to be constant throughout the sensible range of the several sensory continuua, and recently Stevens (1957, 1958b) has presented data suggesting that jnd's are not equal in size. Luce and Edwards (1958), meanwhile, have pointed up certain weaknesses in Fechner's mathematical reasoning. The most serious challenge, however, resides in Stevens' (1961) argument that it is not only empirically unjustified but logically untenable to generate scales from measures of confusion. "Starting from scales", he states (1961, p. 37), "we can determine error distributions, but starting from assumed error distributions we cannot establish scales." To replace the Weber-Fechner law and the law of comparative judgment, he presents the power law.

Using direct estimation methods in which the subject is required to identify his judgment in terms of a numerical scale referenced to an arbitrary value assigned some reference stimulus (e.g., the effective threshold), Stevens (1960) has assembled data on more than twenty sensory dimensions from which he has constructed rectilinear log-log psychophysical functions with slopes varying between .3 and 3.5. These exponents identify the power relationship in the power law:

$$\text{Judgment} = C\ (\text{Judged Stimulus} - \text{Effective Threshold})^n$$

where C is a constant having to do with the magnitude of the numbers used in the response scale and n is the exponent of the power function.

Stevens has concluded that the power law allows the construction of ratio scales. However, this conclusion, as Rosner points out (1962), needs qualification. On physical ratio scales, both intervals and ratios are defined. In addition, intervals and ratios are linearly related. If Stevens' scales are true ratio scales, interval judgments would be a linear function of ratio judgment. But as Stevens himself has shown (Stevens and Galanter, 1957), category scales are concave downward when plotted against magnitude estimation scales. Furthermore, recent data (Black and Bevan, 1960; Bevan and Pritchard, 1962) which demonstrate that subliminal stimulation may induce shifts in category scale values raise doubt concerning the absolute magnitude of Stevens' reference values.

Garner, Hake, and Eriksen (1956) have recently suggested that in direct magnitude estimations the subject is actually matching the perceived stimulus magnitude with subjective impressions of number, and if the psychophysical function for number were taken into account, the difference between equal-interval (psychophysical category scales are generally treated as equal-interval scales) and direct-magnitude scales would be resolved. If one considers the dimensions of sensory intensity and subjective number both to be special cases of a general scale of subjective magnitude, then a subjective number becomes the reference stimulus against which the judgments of magnitude are scaled. Subjective values for numbers are, we know, not fixed, but depend upon the number context in which they occur. Viewed in this light, the number reference value becomes a part of the judgmental background specified by adaptation-level theory.

According to this latter point of view (cf. e.g. Helson, 1959), the judgmental reference value is the product of both what Rosner refers to as first-order (stimulus) and second-order (contextual) independent variables. Formally defined, it is as follows:

$$AL = \bar{S}^p, B^q, R^r$$

Where A.L. is the reference value, $\bar{S}$ is an average of the stimuli being judged, B is the present background stimulation against which judgment occurs, and R is residual stimulation, i.e., those aspects of context contributed by past experience and other variables not formally identified in the psychophysical experiment. In many psychophysical experiments the experimental situation is such that the residual is effectively zero. For such instances, the expression may be written: $AL = \bar{S}^p B^q$. Since individual judgments are referenced to adaptation-level, they may be described as follows:

$$\text{Judgment} = S_i - AL$$

where S_i is the magnitude of the stimulation processes associated with the stimulus being judged.

If it is assumed that there is some general constant that defines the relationship between the physical and the subjective scale involved,

$$\text{Judgment} = (S_i - AL)^n.$$

When the power law is written as the general case, it is as follows:

$$\text{Judgment} = S_i^n.$$

When it is recognized that scalar magnitude is referenced to the effective threshold, it becomes:

$$\text{Judgment} = (S_i - S_o)^n.$$

Now, if the effective threshold is identified with the subjective neutral point rather than the absolute threshold and thus rendered equivalent to A.L., the expression becomes:

$$\text{Judgment} = (S_i - AL)^n,$$

and the power law is seen to be a special case of the adaptation-level principle, derived from situations in which the contribution of B and R variables is usually relatively small.

All of the approaches to quantification reviewed here have involved the same measurement model, the single linear dimension of classical physics. There is currently much concern with the fact that in many judgmental situations, stimuli to be judged vary simultaneously on several dimensions. This has led to two kinds of work: (1) empirical studies in which a limited number of dimensions are systematically covaried and the effect of this upon judgment ascertained (Turner and Bevan, 1962; Bevan and Turner, 1962), and (2) the elaboration of scaling theory to accommodate sets of stimuli which vary on an unknown number of dimensions. In this latter instance, stimuli are represented as points in Euclidian space, the similarity of any two being indicated by the distance they are apart in this space (cf. e.g. Messick, 1956; Torgerson, 1958).

The Materials of the Experimental Psychologist

The goal of the experimental psychologist is to establish scientific facts. These differ from other kinds of facts -- legal facts, theological facts, etc. -- in that they result from the application of methods which differ in certain important ways from those used in the non-scientific disciplines. But most, if not all, facts have one thing in common: they are derived, on the one hand, from concepts and on the other from data. "All that is factual," Goethe is said to have observed, "is already theory". Facts take shape against a conceptual background. At the same time, new concepts grow out of already established fact (Bevan, 1953).

On the use of models. The conceptual framework of the experimental psychologist is many-faceted. Certain assumptions, as I have noted above, are related to the general nature of psychological phenomena. Others are related to the nature of psychology as an intellectual discipline and the methods appropriate to its pursuit. On these are overlaid the particular theoretical bias with which the investigator feels congenial. Inspection of the activities of the experimental worker as he formulates a problem prior to conducting an experiment and as he prepares to interpret his data after this experiment is completed reveals a frequent preoccupation with analogical reasoning. This helps him to see the nature of the relationships he anticipates among his data before they are collected and to tie them together with other data after being collected. Much has been written in recent years about the place of models in the logic of science. Braithwaite (1953), for example, has commented at length concerning the relationship of the model to the corpus of a theory. Von Bertalanffy (1952) has classified models into contrasting types according to whether they are static or dynamic, molecular or molar, and material or formal. And in addition, Lachman (1960) has differentiated four types of modular function: as a mode of representation, as rules of inference, as interpretations of the calculi of theories, and as pictorial representations.

In addition, distinction must be made between the logician's concern for the formal properties of models and the practicing experimentor's use of models as instruments of science-making. The latter typically will not (nor need not) give any more attention to the logical properties of the model than he will to the internal workings of a piece of equipment that he uses in data collection. Of course, he must discover when it is appropriate to use the model, and when a disanalogy exists; that is, when the model differs in some important respect from the empirical relationships obtaining among the variables of his experiment. But this last is a matter of common sense, experience, intuitive skill, conceptual artistry -- things which are not the necessary consequences of preoccupation with logical form and structure. Indeed, it seems highly probable that the practicing experimentor's use of models is, in principle, more simple and direct than logical analysis might lead one to expect. Models must be simple and direct to be effective and subtlety exists not in their structure but, as Oppenheimer put is (1956), in finding the disanalogy that will enable the scientist to preserve what is right about his analogy. Each of the laws of Newtonian mechanics may be extended to atomic mechanics if it is assumed that the momentum and the coordinate are not numbers but objects such that the product of the momentum and the coordinate is not the same as the product of the coordinate and the momentum, but that the difference between these two products is an imaginary, universal, atomic constant (Oppenheimer, 1956).

Too much cannot be said for recognizing the importance of common sense to the activity of model-making. While it is true that the instruments of a science make possible concepts and data that transcend the realm of ordinary experience, this does not mean that what we learn of nature from ordinary experience lies outside our science. Oppenheimer (1956) describes a number of ideas identified with common sense that have been incorporated into modern physics after 300 years of rejection from physical doctrine: the physical world is not completely determinate; there is a limit to what can be objectified without reference to the actual operations of observations; the phenomena being studied are inseparably linked to the methods of study in the evolution of data; events are indivisible, individual, and, in their essentials, not reproducible.

Finally, it is also important to recognize that success in the formalization of a science depends upon its level of development. Regardless of the methodological sophistication of its adherents, one cannot expect the same degree of conceptual articulation in a field with few firmly established facts or general principles as one finds in a more advanced field. Certainly, to transpose a model from one science to another when the relationships between the two sciences are not clearly understood is an exercise of indeterminate merit. The fact that most discussions of the problems of theory construction and other matters in the philosophy of science which are written by psychologists draw heavily upon physics for illustrative material rather than upon psychology itself prompts the suspicion that a preoccupation with issues of so highly systematic a nature may be somewhat premature.

On the nature of data. As was noted earlier, there is an important difference between observations and data. Data are statements that describe observational events. They may be expressed in many forms -- in the natural language of every-day speech, in the technical language of some particular intellectual discipline, in graphic form, and in terms of one or another form of mathematics. However, regardless of their form of expression, all data have one thing in common: They symbolically represent, they do not reproduce, observations. Different forms of data communicate information from observations with different degrees of efficiency. The non-technical verbal forms of everyday life, while they provide the richest vocabulary for the description of experience, at the same time are the most imprecise. In contrast, the formal language of a mathematical system, while its range of expression may be comparatively limited, will convey information with a minimum of excess meaning. None will completely describe an observation which the observer regards to be important or otherwise relevant. This is further complicated by the apparent fact that observers differ in the sensitivity and the comprehensiveness of their observations.

Psychological data fall into two broad classes: those having to do with the behaving organism and those having to do with the situation in which behavior occurs. Situational data, in turn, are of two general types: stimulus data, those related to events impinging upon the organism, and response data, those describing the organism's reaction to stimuli. The traditionally-defined experimental psychologist (perhaps because the classical psychophysicist has for so long viewed the psychophysical problem as one of quantifying the correspondence between physically defined stimuli and sensation and perhaps because the classical behaviorist for so long insisted upon equating objectivity with the language of physical science) has directed his attention for the most part to situational variables. Relatively little interest over the years has been evidenced in laboratory studies of behavioral development, of individual differences, of behavioral genetics, etc. By the same token, little formal concern in experimental design has been evidenced for variables like pre-test experience and differences in temperament. Meanwhile, it is becoming increasingly clear to the experimental psychologist that it is possible to perform perfectly good experiments without being committed to the classical S-R epistomology, that the biggest difference in the laboratory work of the physicalist and the phenomenalist may be one of precision, and that quantification is not limited to the dimensions of physics. Thus, one may in principle investigate curiosity as readily as hunger and measure attractiveness as easily as brightness.

Indeed, if one examines the laboratory practice rather than the metaphysics of the experimental psychologist, one will soon discover that purely physical knowledge has never been adequate to the identification of stimuli and responses. To begin with, stimuli are identified by their capacity to evoke particular kinds of responses, and the experimenter, in designing an experiment, falls back upon his own experience as an observer in the selection of S-R variables to investigate. This is so whether the subjects are human or subhuman. I strongly suspect that even the most hard-headed behavioristic purist selects as reinforcement for his rats in the lever-pressing or T-maze situation objects which he believes are meaningful to them, though he could never admit this in his theory. But this last is exactly what is necessary. Theoretical exercise must be made consistent with laboratory practice, and if the identification of S and R reflects, to use Koch's phrase (1959), "the perceptual sensitivities of human observers", this property must be incorporated into their definition within the _corpus_ of a theory.

On converging operations. Two broad methodological problems are faced in the definition of stimuli and responses: those of validity and reliability. In the case of the former, concern is that the measures or other operations that the experimenter devises to represent a variable under consideration are truly representative of it. In the latter, interest is in the consistency resident in the measure. For the past 30 years psychologists have been enthusiastic advocates of the doctrine of operationism and in this time the literature reveals many instances of its uncritical application. Investigators appear to proceed to their experiments with the assumption that if they can invent an operation to measure a concept it will be adequate to doing so. Thus, it may be taken for granted that if a rat is deprived of food for a certain period of time he will be motivated to run through a maze to obtain a pellet of food in the goal box or if a human subject is given a short time-interval to complete an experimental task he will be working under stress. Recently, many American pharmaceutical houses set up operant conditioning laboratories for the screening of compounds without, I am sure, any clear notion of the significance of such screening data.

Whether or not a concept measured by different sets of operations is one and the same (i.e., whether length measured by a meter stick is the same as length measured by triangulation) is a matter to be settled by the philosophers of science. Meanwhile, Garner, Hake and Eriksen (1956) point out that Bridgman, in his formulation of the principle of operational definition, was talking about sets of operations, and they propose that confidence in the validity of definition may be enhanced by the planning of sets of interrelated experiments which provide complementary data on variables under consideration. These they call converging operations. Their discussion of perceptual defense is instructive. If, for example, recognition times are obtained for vulgar and neutral words and the former are longer than the latter, this difference may be attributed to perceptual defense. But it may also be due to response suppression. In order to discriminate between these two hypotheses, an experimenter might add a condition to his experiment such that vulgar stimuli were used to evoke non-vulgar responses and vice versa. Not all multiple-operation experiments are converging. Garner, Hake and Eriksen (1956) indicate two types which are not. (1) Repeat operations may enhance the reliability of the original finding but they do not clarify its nature. (2) Similarly transform operations in which one set of conditions is simply substituted for the original (e.g., synonyms used in place of the original vulgar and non-vulgar words) can do no more than reinforce the original conclusion. An impressive illustration of the value of converging operations is seen in Muenzinger's series of experiments on the nature of punishment to be referred to later.

The problem of reliability is more competently understood by the experimental psychologist and requires little discussion here. The experimenter has a variety of devices available to aid in achieving maximum reliability. He may use pointer-readings rather than verbal descriptions, he may use more refined equipment, he may employ a variety of procedural controls such as counterbalancing to reduce systematic effects associated with incidental variables, and he may use statistical procedures to segregate the incidental sources of variance once these get into his data. Both validity and reliability are important concerns, but the former must take precedence over the latter. Nothing is gained by niceties of measurement if we are not quite sure of what we are measuring or whether or not it is, in fact, worth measuring.

The Methods of the Experimental Psychologist

There are at least two ways in which a short section on the methodology of experimental psychology can be written: one may provide a compendium of procedures, or one may attempt to identify the attitudes and assumptions the experimental psychologist holds concerning his methods. We prefer the latter.

Unlike many other young sciences, psychology, from its earliest period, has emphasized the role of experimentation as the basic method for carrying on its affairs. Unlike biology, with which it is often identified, it has had in its history no taxonomy and no periods of natural history. With perhaps the exception of the nosological systems of the early abnormal psychology, there has been no preoccupation with classification of behavioral phenomena, and, perhaps with the exception of certain recent publications in human engineering, there have been few if any handbooks of the sort identified with physical science and engineering. The emphasis upon the formal experiment appears to be tied to the persistent demand for quantification. It is interesting to note that, in his Handbook of Human Physiology, Johannes Müller proclaimed that the speed of the nerve impulse defied measurement and his treatment of sensory processes and the mind were largely metaphysical and speculative. In little more than a decade after Müller's pronouncement, Helmholtz had measured the speed of the nerve impulse and had found it, indeed, to be rather slow. Still a few years later, Fechner had proposed his solution to the mind-body problem and had provided the methods by which to attain solution.

Orienting attitudes. One implicit attitude concerning the significance of experiments is particularly widespread among experimental psychologists. They are prone to think of their results in extremely general terms. This generality involves both species and situations. Except as he is, from time to time, caught short in his application of experimental data, the experimental psychologist is inclined to view his findings as valid without respect to the readily apparent differences between his laboratory situations and those obtaining outside the laboratory. Indeed, he is likely not to keep in mind the great dependence that behavioral data have upon the particular methods used in securing them. One need only to read the usual textbook treatment of specific behavioral phenomena to verify this practice. The same state of affairs holds for species. In the development of particular theoretical issues, data from a variety of sources are brought together in the supporting evidence and one senses an equation of the white rat with the human observer. That is, there appears to be the assumption that the white rat is substitutive for the human observer in the explanation of behavior in the general case. From this point it is an easily-taken step to the attitude that a white rat is a human observer, albeit a rather simple one, in some laboratory situation under consideration.

In recent years, Brunswick (1956) has made an eloquent statement of the principle that experiments must be representative in design: that is, that they must be so designed that they will yield data which will correlate positively with data from the situations to which one intends to generalize his experimental data. Representative experiments are identified by two criteria. (1) They have ecological validity, that is, they sample a range of situations that are typical for the type of subject being observed; and (2) they are functional in intent, that is, they sample independent-dependent variable relationships over a typical range of magnitudes. It is paradoxical that attention in psychological experiments, outside of psychophysics, has been directed more to the identification of independent variables, less to the determination of S-R functions. Even with an increased use in the past several decades of complex designs which facilitate the determination of such functions, greater attention persists toward the identification of variables and their interactions.

Psychological experiments also have been concerned more often with situational in contrast to intraorganismic determinants of behavior. This, perhaps, is because the former are more accessible, more easily identified, and more readily controlled.

Psychological experiments generally are simple in conceptualization. Though complex design procedures have been introduced, most variables are present in, at most, one or two degrees of magnitude, and interactions above the first-order are difficult to interpret. A major virtue of complex designs, of course, is that they allow the reduction in the size of error terms and thus the demonstration of effects with smaller samples of subjects. Moreover, the indices that constitute data in psychological experiments are simple ones. The response operations required of the subject generally are simple acts like saying "yes" and "no", pressing a key, or tracking a particular signal.

Performance is scored for the frequency with which responses or response errors occur, their latency and/or duration, and only relatively infrequently involves direct estimates of magnitude. When physiological measures are used as indices of behavior, these are more frequently than not simple measures.

Error in the experimental psychologist's mind is something to be eliminated. His greatest concern is with variable error and its effect upon the precision of his data. He will repeat observations to reduce error; he will use counterbalancing and other design procedures to reduce its bias; and he will employ complex designs to remove the influence of known variables and their interactions from the assessment of the effects in which he is primarily interested. The constant errors typically have held little interest for him, and the specialty of individual differences has never been a major area of psychological investigation.

Levels of measurement. There are basically four levels of measurement (Stevens, 1958a). That is, the numerals assigned to represented observations will differ in the information they convey and these differences may be ordered into four categories. In the very least, they can simply identify the observational event or class of events. Or they may indicate position in a linear array of events. Examples of these two levels are the seat numbers assigned to students in a class and the ranking of scores on an essay quiz. Often evaluative criteria are complex and it is not possible, as Coombs (1953) has pointed out, to achieve a pure case of rank-ordering. Then one must be content with a partially-ordered metric. For example, if the answers on an essay test are evaluated for both factual accuracy and original thought, Test Paper A may be clearly superior to Test Paper B on both and the two may be ordered so that A is higher than B. Meanwhile, Test Paper C may be factually less accurate than B but display more originality. Thus it is not clear whether B is higher than C or C is higher than B. Unless some *a priori* rule has been made about the relative weighting of the two factors or a decision to score the tests independently on the two criteria, the strongest form of scale that can be achieved is the partially-ordered scale.

Most frequently, the data of the experimental psychologist are of the ordered type and the number of analytical operations he can perform legitimately upon them is limited to three: the computation of medians, percentiles, and rank-order correlations. Meanwhile, his goal in measurement is to attain strong scales. These are of two types: interval and ratio, which respectively identify position on a linear dimension (providing information about scalar distance between magnitudes may be scales) and allow the precise estimate of relative magnitude. Examples of interval and ratio measures are the numerical scores on a standardized achievement test and the time-scores on tests of manual dexterity. Stevens (1958a) has maintained that certain sensory scales like the sone and the bril scales are ratio scales. For a scale to be a true ratio scale it must have a true zero. Stevens has identified this as the absolute threshold (more recently, he has used the vague term, *effective threshold*). However, recent work demonstrating the predominant influence of contextual factors, particularly subliminal stimulus inputs (Black and Bevan, 1960), brings into question the assertion of sensory ratio scales.

The interval scale is perhaps all-around the most useful and the experimenter will do all that he can to achieve it, often by the simple expedient of making certain heuristic assumptions and intuitively treating his scale as one of equal-appearing intervals. However, often it is not possible to even assume a pure interval scale, at which time the ordered metric scale may be appropriate. In this form of scale, information on relative distance between items at different parts of the scale is available although none is provided on the size of the scale unit of measure. Coombs (1950) provides a method for the construction of the partially-ordered metric in his unfolding technique.

<u>The psychophysical methods</u>. Psychophysics is, as we noted earlier, the prototype of strong scaling. The psychophysical methods may be described as being either indirect or direct scaling methods.

The indirect methods are identified with Fechner and Thurstone and are based on the assumption that sensations may not be measured directly. Fechner's expedient was to assume that the absolute threshold constituted the limiting value on a sensory dimension and that jnd's were equal. He thus constructed psychophysical functions by means of the operation of cumulating jnd's. Similarly, Thurstone proposed to infer psychological magnitude from a knowledge of the proportions of times each stimulus is judged greater or less than every other item with which it is compared.

In contrast, Stevens has assumed an isomorphism between the operations of response and the sensory magnitudes underlying responses. He thus has proceeded to construct scales directly from the experimental operations of bisection, fractionation, and, most recently, magnitude estimation.

Reference has already been made to the shortcomings of Fechner's procedure of cumulating jnd's, and at present, no convincing experimental check on the validity of the Law of Comparative Judgment exists. Nor is the record of the direct scaling methods any better. Garner, Hake and Erickson (1956), in their discussion of converging operations, comment that there is no evidence that sensory magnitude scales are true reflections of a sensory process. Stevens' scale for the direct estimation of loudness (1956) is identified for special comment. When a subject is required to communicate his impressions of loudness in terms of a numerical scale, he may be matching his impressions of loudness with his impressions of numerical size rather than reflecting a metric property of the perceptual system.

Another way of classifying the psychophysical methods has been proposed recently by Rosner (1962). This is based on the type of computational devices used in arriving at numerical values. There are what Rosner calls S statistics and R statistics. The methods that employ S statistics are the methods used in threshold determination and the adjustment methods. These are the indirect methods just described. The computational routines associated with the S methods describe the distribution of a particular response with reference to the sample of stimuli that arouse it. The psychophysical function is constructed by plotting the corresponding conditional probabilities or median adjustment against the scaled stimulus values. The R methods involve summarizing the distribution of responses to particular stimuli. Scale values derive either from the conditional probabilities or the average magnitudes assigned to the several stimuli. Thurstone and Stevens' scaling methods are both examples of the R-methods as is the use of reaction time data (Kellogg, 1931) in scaling.

The methods of multidimensional scaling developed within the last decade provide for the scaling of similarity among stimuli, when the dimensions of similarity are unknown, in terms of distance between points in a Euclidean space of least dimensionality. They are therefore R methods. Another approach to the problem of multidimensionality is worth exploring. This involves the application of category scaling to situations in which the independent variables are some limited and specified number manipulated in fully prescribed ways.

For example, in a recent experiment Bevan and Turner (1962) obtained judgments of size when the variable of lightness was allowed to co-vary with size in a perfectly positive, negative, or random fashion. From these data three psychophysical functions were constructed. Using the condition of random correlation as the basis for comparison, it was found that the judgment of size was augmented in the condition of positive correlation and that the opposite held true for the case of negative correlation. Next, by fitting equations to these three functions and deriving a numerical estimate of their slopes, it was possible to construct a higher-order function from which to predict a psychophysical function for size estimation when, everything else being equal, any degree of relationship existed between this variable and stimulus lightness.

Psychology has taken its quantitative models primarily from the physical sciences. An approach worth examining is that of the similarity and dimensional methods used during the last thirty or more years for the purpose of comparing empirical functions with simple quantitative models of biological systems. In a recent review article, Stahl (1962) identifies a dozen physiological constants deducible from the allometric equation of Huxley. Von Bekesy's work with models of the inner ear (1960) provides an example of interest to experimental psychologists.

The Art of Experimental Psychology

The skillful experimental psychologist is a person who possesses a thorough knowledge of the experimental literature, particularly that segment related to his area of special interest; who is well trained in the techniques of apparatus construction, maintenance, and use; who is well versed in the methods of experimental design, observational procedure, and statistical analysis; and who is informed about the logic of moden science.

But he must be more than this. He must be thoroughly conversant with the lore of experimental psychology. He must possess a creative imagination that will allow him to translate hypotheses into effective experiments and enable him to detect at the earliest possible time the unexpected in his data and capitalize upon it for further lines of investigation. Being familiar with the lore of a field involves knowing the myriad little things that contribute to effective experiments. This is knowledge that is not taught in formal classes or presented in textbooks and handbooks. It largely is omitted from experimental reports. It is acquired through conversations and collaboration with other investigators and through doing experiments oneself. It consists of a knowledge of endless detail that makes the difference between a successful and an unsuccessful experiment: what type of design is best for a particular problem; what sorts of and how many subjects are needed; what instructions are best and how these are best given; what incidental factors in the situation may be ignored and what must be controlled; what responses need be observed and recorded; what sorts of data are best, etc.

What lies behind creative imagination we cannot at present specify with confidence. But we do know that there are persons who can do exceedingly clever experiments that cut to the heart of an issue while others, just as intelligent and just as well-trained in the facts and methods of psychology, seem to lack the knack. Guthrie (1959), in a discussion of the orienting attitudes of the experimental investigator of learning, points out that patterns of stimuli and patterns of attendant responses have a psychological significance that must be dealt with. "It is not enough," he says (1959, p.165) "that they be available in the physical situation nor is it enough that the organism's attention orients sense organs to receive them; it is further necessary that they have meaning for the responding organism." Psychologists who plan conditioning experiments undoubtedly choose their stimuli in the light of their past experience that these do get responded to by the subject. I would go a step farther and suggest that the skillful experimenter has the capacity of placing himself in the role of subject and viewing the experimental situation from this vantage point. In performing this exercise, he will know not only whether or not his chosen signals will evoke the response he is interested in, but he also may gain some insight into the subtleties of the situation as the subject perceives it and therefore some comprehension of the behavior that he will observe in his experiment.

An experimenter at work. I should like to conclude this discussion by describing a series of experiments by Muenzinger and his colleagues on the properties of punishment and the validity of the Law of Effect. They are reported in a set of nine papers, entitled "Motivation in Learning", that appeared between 1931 and 1941 and were summarized by Muenzinger in the 1945 Annual Research lecture presented at the University of Colorado (1946). I have chosen this series, not because I have any special interest in the problem they were devised to resolve or because I find Muenzinger's theoretical orientation particularly appealing. Rather, I have selected this program of studies because the experiments are simple in conceptualization and because they illustrate very nicely the mind of an experimenter at work. They demonstrate that no experiment, no matter how comprehensive in design, fully resolves a theoretical issue or establishes an experimental hypothesis. What each does accomplish, if it is well-executed, is added clarity and an increased articulation of the question that gave rise to the study in the first place.

Muenzinger's goal was to clarify the relationship of punishment to performance efficiency. In all of his experiments, his vehicle was the single-unit T-maze and the subject's task was that of choosing a lighted alley leading to food in the goal box. Albino rats can reach errorless performance in this task in an average of about 100 trials. If shock is added for incorrect choices, the same level of accomplishment will be reached in about 40 trials. Thorndike's Law of Effect would account for such facilitation by stating that the annoying consequences of entering the incorrect alley produced a weakening of the S-R bonds that linked the stimulus cues provided by the alley to the entering response, while the satisfying consequence of entering the correct alley strengthened this S-R connection.

But, as Muenzinger realized, the Law of Effect describes only one of several possible reinforcement patterns. Therefore, his first concern was with the effect upon performance of shock for correct responses. He predicted that if the correct response were both rewarded and punished, performance would be generally poor, since the strengthening effects of the former would be counteracted by the latter. In his basic experiment he ran three groups of rats: (1) which received only food on the correct trials; (2) which received food on the correct and shock on the incorrect trials; and (3) which received both food and shock for the correct response. The results were quite surprising. The group receiving food alone reached criterion (two consecutive days of 10 errorless trials each) in 114 trials as expected. The group rewarded for correct and punished on incorrect trials took 39 trials as expected. But the group both rewarded and punished on the correct trials averaged only 49 trials and were not significantly poorer than Group 2.

In attempting to account for the facilitating effect produced by punishing the correct response, Muenzinger reasoned that punishment might influence performance efficiency not by weakening incorrect S-R bonds but by making the animal more alert to the problem's critical cues. Therefore two more groups of animals were tested: (4) which received shock both for the correct and incorrect alleys; and (5) which was shocked in the starting alley prior to reaching the choice point. The group which was shocked for both correct and incorrect responses reached criterion in an average of 40 trials, equaling the performance of Groups 2 and 3. In contrast, the group shocked in the starting alley took 118 trials to reach criterion, indicating that shock under these conditions had no facilitating effect at all. In order to further check on the possibility that punishment exerted its influence through general arousal, a sixth group received shock in the starting alley and in both correct and incorrect alleys and met criterion in 49 trials, a level of performance not significantly different from that of Groups 2, 3, and 4.

Muenzinger's conclusion at this point was that punishment increased performance efficiency by sensitizing the subject to the cues critical to solution of the task. He, furthermore, hypothesized that reward and punishment are complementary states related to the attainment or frustration of goals. Thus securing food and escaping shock both may be rewarding, while not securing food and receiving shock both may be punishing. To check the possibility that escape from shock is rewarding, Group 7 received shock from start box to goal box with no food in the goal box. Its mean criterion performance was 52 trials, not significantly different from 2, 3, 4 and 6 and clearly better than Group 1, the food-alone group.

Muenzinger next turned to examine another aspect of the Law of Effect. This hypothesized not only that punishment produced its effect by weakening S-R bonds but that this resulted from the annoyance associated with it. If, as Muenzinger had concluded, punishment was effective because it alerted the subject to important cues, then it was necessary to ask if this alertness resulted from annoyance or from some other cause. This meant replacing shock at the choice point with some other type of reinforcement. Group 8 received a buzzer in the correct alley; Group 9 received it in the incorrect alley. For Group 10, the grids were removed from both alleys and the animal was faced with jumping a gap to reach the goal box. For Group 11 the gap was located in the starting alley prior to the choice-point. The substitution of the buzzer for the shock in Groups 8 and 9 produced no facilitation of performance. In contrast, the group with the gap in both alleys averaged 49 trials to criterion; its control, Group 11, 110 trials. Delay at the choicepoint was as effective as shock at the choicepoint in facilitating learning. In order to get some clearer notion of why this was so, Muenzinger next examined other data he had available on all his groups. In groups with scores above 100 the average amount of V.T.E. behavior (head-turning at the choice-point) was only about one half that observed in the groups performing at a level of 40-50 trials to criterion. He, therefore, concluded that any condition, including punishment, that will cause the subject to pause at the choice-point and indulge in V.T.E. behavior will facilitate performance.

This last hypothesis was tested by observing a group (12) in which the subject was simply restrained in the alley of its choice for 3-5 seconds on each trial by the lowering of transparent doors and then allowed to continue to the goal box. These rats reached criterion in 68 trials, and were considerably better than the food-alone group (1).

Muenzinger's interpretation of the role of punishment in learning, based on the performance of these 12 groups of rats in the T-maze, is that it facilitates performance because it prompts the subject to pause at the choice-point and this pause increases the probability that he will detect the cues that must be discriminated for successful solution of the experimental task.

References

Adams, D. K. Learning and explanation. In Learning theory, personality theory, and clinical research, the Kentucky symposium. New York: Wiley, 1954, 68-80.

Allport, G. W. Scientific models and human morals. Psychol. Rev., 1947, 54, 182-192.

Bevan, Jr., W. Moden psychologists: Scientific woozle hunters? Nordisk Psychol. Monografiserie, 1953, No. 4.

Bevan, W. and Pritchard, Joan F. The effect of subliminal tones upon the judgment of loudness. Unpublished manuscript.

Bevan, W. and Turner E. D. The influence of lightness upon the judgment of size. Unpublished manuscript, 1962.

Black, R. W. and Bevan, W. The effect of subliminal shock upon the judged intensity of weak shock. Amer. J. Psychol., 1960, 73, 262-267.

Braithewaite, R. B. Scientific explanation. Cambridge: Cambridge Univ. Press, 1953.

Brunswik, E. Perception and the representative design of psychological experiments. Berkeley: Univ. Calif. Press, 1956.

Campbell, N. R. An account of the principles of measurement and calculation. London: Longmans, 1928.

Carnap, R. The methodological character of theoretical concepts. In Feigl, H. and Scriven, M. (Eds.) Minnesota studies in the philosophy of science, Vol. I. Minneapolis, Minn.: U. of Minn. Press, 1956, 38-76.

Coombs, C. H. Psychological scaling without a unit of measurement. Psychol. Rev., 1950, 57, 145-158.

Coombs, C. H. The theory and methods of social measurement. In Festinger, L. and Katz, D. (Eds.) Research methods in the behavioral sciences. New York: Dryden Press, 1953, Ch. II.

Garner, W. R., Hake, H. W. and Eriksen, C. W. Operationism and the concept of perception. Psychol. Rev., 1956, 63, 149-159.

Gibson, J. J. The perception of the visual world. New York: Houghton-Mifflin, 1950.

Guilford, J. P. Psychometric methods, 2nd Ed. New York: McGraw Hill, 1954, Ch. 2.

Guthrie, E. R. Association by contiguity. In Koch, S. (Ed.) Psychology: a study of a science, Vol. 2. New York: McGraw Hill, 1959.

Hebb, D. O. The organization of behavior. New York: Wiley, 1949.

Hebb, D. O. The role of neurological ideas in psychology. In Krech, D. and Klein, G. S. (Eds.) Theoretical models and personality theory. Durham, N. C.: Duke Univ. Press, 1952, 39-55.

Helson, H. Adaptation level theory. In Koch, S. (Ed.) Psychology: a study of a science. Vol. I. New York: McGraw Hill, 1959. 565-621.

Johnson, H. M. A simpler principle of explanation of imaginative and ideational behavior and of learning. J. comp. Psychol., 1927, 7, 187-235.

Katz, J. J. and Halstead, W. Protein organization and mental function. In Halstead, W. (Ed.) Brain and Behavior. Comp. Psychol. Monogr., 1950, 1-39.

Kellogg, W. N. The time of judgment in psychometric measures. Amer. J. Psychol., 1931, 46, 65-86.

Koch, S. Toward an indigenous methodology. Address to Eastern Psychological Association, April, 1959.

Koch, S. Psychological science versus the science-humanism antimony: intimations of a significant science of man. Amer. Psychol., 1961, 16, 629-639.

Lachman, R. The model in theory construction. Psychol. Rev., 1960, 67, 113-12?.

Luce, R. D. and Edwards, W. The derivation of subjective scales of just noticeable differences. Psychol Rev., 1958, 65, 222-237.

McGuigan, J. The experimenter -- a neglected stimulus object. Psychol. Bull.. In press, 1962.

Messick, S. J. Some recent theoretical developments in multidimensional scaling. Educ. psychol. Meas., 1956, 16, 82-100.

Muenzinger, K. F. Reward and punishment. Univ. Colo. Stud., Gen. Series (A), 1946, 27, (No. 4), 1-16.

Oppenheimer, R. Analogy in science. Amer. Psychologist, 1956, 11, 127-135.

Popper, K. R. The logic of scientific discovery. New York: Basic books, 1959, Ch. 7.

Rosner, B. S. Psychophysics and neurophysiology. In Koch, S. (Ed.) Psychology: A study of a science, Vol. 4. New York: McGraw Hill, 1962, 280-333.

Skinner, B. F. The behavior of organisms. New York: Appleton-Century Crofts, 1938.

Stahl, W. R. Similarity and dimensional methods in biology. Science, 1962, 137, 205-212.

Stevens, S. S. The direct estimation of sensory magnitudes -- loudness. Amer. J. Psychol., 1956, 69, 1-25.

Stevens, S. S. On the psychophysical law. Psychol. Rev., 1957, 64, 153-181.

Stevens, S. S. Measurement and man. Science, 1958, 127, 383-389.

Stevens, S. S. Problems and methods of psychophysics. Psychol. Bull., 1958, 54, 177-196.

Stevens, S. S. The psychophysics of sensory function. Amer. Sci., 1960, 48, 226-253.

Stevens, S. S. Toward a resolution of the Fechner-Thurstone legacy. Psychometrika, 1961, 26, 35-47.

Stevens, S. S. and Galanter, E. H. Ratio scales and category scales for a dozen perceptual continua. J. exp. Psychol., 1957, 54, 377-411.

Thurstone, L. L. Psychophysical analysis. Amer. J. Psychol., 1927, 38, 368-389.

Thurstone, L. L. A law of comparative judgment. Psychol. Rev., 1927, 34, 273-286.

Torgerson, W. S. Theory and methods of scaling. New York: Wiley, 1958, Ch. 11.

Turner, E. D. and Bevan, W. The simultaneous induction of multiple anchor effects in the judgment of form. J. exp. Psychol., in press.

Von Békésy, G. Experiments in hearing. New York: McGraw Hill, 1960.

Von Bertalanffy, L. Theoretical models in biology and psychology. In Krech, D. and Klein, G. S. (Eds.) Theoretical models and personality theory. Durham, N. C.: Duke Univ. Press, 1952, 24-38.

Zener, K. The significance of experience of the individual for the science of psychology. In Feigl, H., Scriven, M., and Maxwell, G. (Eds.) Minnesota studies in the philosophy of science, Vol. II. Minneapolis, Minn.: Univ. Minn. Press, 1958, 354-369.

Technical Report No. 8, 1962

Studies of Anomalous Contrast and Assimilation*

Harry Helson
Kansas State University

* This report describes work supported in part by contract Nonr-3634(01) between Kansas State University and the Physiological Psychology Branch, Office of Naval Research.

The present paper was presented at the spring, 1962 meetings of the Armed Forces - NRC Committee on Vision held in Washington, D. C. It will appear in the December, 1962 issue of the *Journal of the Optical Society of America*.

November, 1962

Introduction

During the past few years I have been concerned with anomalous contrast and assimilation effects such as the von Bezold "spread" effect and the Gelb spotlight phenomenon in the hope of obtaining quantitative data which would make it possible to envisage these and similar phenomena in terms of known visual mechanisms. In the von Bezold effect[1] chromatic areas with white arabesques appear lighter than with black designs, and in the Gelb phenomenon a spinning black disk illuminated by a cone of light in a dimly lit or dark room with black walls appears very light gray or white until a small white surface is seen in front of the disk whereupon it turns very dark gray or black. The Gelb effect[2] has been explained by such molar concepts as "noticing the illumination," "belongingness," "appurtenance," "structure of the visual field," or in terms of "differences between film and surface colors." The rejection of contrast as an explanation of the Gelb effect was based on the supposition that the small white area could not cause such a large change in the appearance of the disc by contrast but no experimental evidence was offered to support this argument. Similarly, the von Bezold spread effect (called 'assimilation' here) has stood opposed to classical contrast. With no quantitative data there has been, until recently, no basis for predicting when assimilation or contrast will be perceived under given conditions. Three studies of assimilation and one study of the Gelb phenomenon are reviewed which show that they are amenable to quantitative treatment, and a theory of their retinal origin is tentatively proposed.

Assimilation and Contrast

Materials and Procedure. The studies of assimilation were made with straight, ruled black and white lines in place of the von Bezold designs, thus making possible quantitative variations in line thickness and line separations (or intervening background area) and nonselective surfaces of high (near-white), low (very dark gray), and intermediate (middle-gray) reflectances were employed in place of chromatic backgrounds. A combination of the methods of paired comparisons and limits was used in the observations. Observers judged the lightness of the two halves of 7x11-in. cards ruled with white and black lines with instructions to compare the lightness of the intervening gray areas. In the first study with Rohles,[3] only lines of 1-mm width were employed with separations ranging from 3-55 mm in steps of 4 mm, and background of 36% reflectance (Munsell value = 6.5). In the second study with Joy,[4] both linewidth and line separation were varied in a 6x6 design in which every width of line was used with every line separation. The linewidths and line separations were 3, 10, 20, 30, 40, and 50 mm. In the third study with Steger,[5] a 5x6 design was employed with linewidths of 1, 3, 5, 11, 17, and 29 mm and line separations of 3, 5, 11, 17, and 29 mm on grays of 14% and 80% reflectance (Munsell values = 4.3 and 9.0). In all three studies the cards were viewed at a distance of 3 m and the judgments of the two halves of the cards were made in terms of a category scale ranging from very, very much lighter through equal to very, very much darker, with numbers from 1-9 assigned them for purposes of computation with 5 standing for equal and numbers above 5 for contrast and those below 5 for assimilation except in the third study where numbers above 5 indicate assimilation. To counterbalance order and practice effects the cards were judged in abba order in the first and third studies, i.e., white-lined sides first, then the dark-lined sides twice, then the white-lined sides, with the observers, and the reverse order, baab, with the remaining observers. In the second study only ab and ba series were employed. Cards were presented in ascending order of magnitude of linewidths and separations and then in descending order of magnitude. Since assimilation and contrast effects are very critical in certain regions, it is necessary to control the area surrounding the stimulus materials. Even objects like white or dark papers or clothes in the visual field of the observers may influence the results, hence the cards were surrounded by large, homogeneous surfaces of identical reflectance. Other details of procedure, evaluation of results, including statistical treatment, etc., will not be given here as they can be found in the original sources.

Results. The reliability of the data is attested by the number of observations in the three studies: in the first study there were 10 Osx4 series x14 stimuli, making 560 observations; in the second study there were 11 Osx2 series x36 stimuli making 792 observations; and in the third study there were 10 Osx4 series x60 stimuli making 2400 observations. The three studies thus embody a total of 3752 observations. In addition, the second study was essentially a replication of an earlier one. Since the second study embodies the first as a special case, only the results of the second and third studies will be discussed in detail in this article.

The data in Table I show that as linewidth increases (reading across the table) assimilation gives way to contrast. In this table numbers above 5 indicate contrast, i.e., the area between the white lines was darker than that between the black lines of the cards. Numbers below 5 indicate that the white-lined half was judged lighter than the black-lined half of the cards and hence that in such cases assimilation occurred. Linewidth of 10 mm is seen to be the dividing point between assimilation and contrast. Contrast increases with linewidth and decreases with line separation, a fact in accordance with the law of diminishing contrast from the borders of contrasting areas, while assimilation is greater with narrow than with wide lines.

Turning to Figs. 1-4 we find consistency in the conditions yielding assimilation and contrast, the two phenomena forming a continuum. Between assimilation and contrast there is a transition zone from one to the other with neither effect present, as we would expect. From Fig. 1 it is seen that lines 3 mm in width yield assimilation with all line separations, the degree of lightening due to the white lines and darkening due to the black lines being roughly inversely proportional to the width of intervening gray background. Additional information is obtained from Fig. 2 in which lightness is plotted against line separation with width of lines as the parameter of the individual curves. With a few exceptions, it is seen that as the width of gray between the lines increases both assimilation and contrast decrease, i.e., the intervening gray tends to be unaffected by the white or black lines. This result bears out the finding by Helson and Rohles that assimilation decreases as line separation increases and it is also in keeping with the classical law of contrast according to which contrast effect decreases with distance from borders.

Effects of width of line and line separation are combined in Fig. 3 wherein lightness is plotted as a function of the ratio of gray width to line width with the former the parameter in the individual curves. For each width of gray, as the ratio of gray to linewidth <u>decreases</u>, contrast increases,--a result readily understandable in terms of classical contrast since the wider the white and black lines relative to the line separations, the greater should be the contrast effect. One important fact emerges from Fig. 3: ratios of white to gray and black to gray are not solely determinative of contrast or assimilation. Thus 1:1 ratios of linewidth to line separation may have very different effects, depending upon the absolute values entering into the ratios: maximal assimilation is found with linewidths of 3 mm and line separations of 3 mm but maximal contrast is found with 50-mm linewidths and 50-mm line separations in this study. The domain of assimilation contains ratios from 1:1 to 16.7:1 under the conditions of this study. The domain of contrast contains ratios ranging from 0.06 to 2.5:1. Hence we must look to visual mechanisms that are affected by absolute as well as by relative amounts of stimulation with the probability that absolute amounts play a greater role than do relative amounts in these phenomena.

The essential unity of contrast and assimilation, hitherto regarded as opposed effects, can be seen from the plot in Fig. 4 wherein the conditions responsible for each effect are mapped for grays of 36% reflectance. While the transition from assimilation to contrast appears abrupt in this plot, actually it comprises a range of conditions. A second region of neutrality, i.e., neither contrast nor assimilation effect, is found beyond line separations of 50 mm where areas are too large to yield either contrast or assimilation under the conditions of this study.

The instability of assimilation-contrast effects under most conditions appears from the steepness of the curves where they cross the neutral axis (5) in Fig. 3 and also in the closeness of the curves for line separations greater than 10 mm, to the neutral axis. Reports in the vicinity of 5 (equal) are near threshold and fluctuate more than do observations made under the more compelling conditions for either contrast or assimilation. For this reason eye movements, point of fixation, and other individual differences in mode of viewing the stimuli may exert pronounced effects on what is seen.

The results of the third study with Steger bear out the previous studies but with some interesting differences. It will be recalled that in this study we used a near-white and a very dark gray background for the purpose of determining what happens as the upper and lower limits of reflectance are approached. With intervening areas of low or high reflectance (14% or 80%) only assimilation in varying degrees is found up to the widest lines and line separations under the conditions of these observations. As seen from Figs. 5 and 6, the former for very light gray background, the latter for very dark gray background, the thinner the lines, the greater is the lightening effect of white lines and the darkening effect of black lines. With wide intervening backgrounds the surfaces tend toward equality of lightness. The bunching of the curves in Fig. 5 throughout the range of line separations and in Fig. 6 at the 17- and 29-mm line separations shows that line separation is the more important of the two variables. This conclusion is borne out with very light and very dark backgrounds by the flatness of the curves shown in Figs. 7 and 8. From these figures it is seen that degree of lightening due to white lines and darkening due to black lines depends more upon the line separations than upon the linewidths: the smaller the line separations, the greater is the degree of assimilation. Conclusions regarding assimilation and contrast effects therefore must take into account not only linewidths and line separations but also the reflectances of the lines and of the intervening areas.

Trigger Contrast (Gelb Effect)

If the sudden, great change in the Gelb disk, illuminated by a spotlight, from 'white' to black as the result of introducing a small patch of white in front of the disk is a contrast effect, then the amount of darkening should be a function of the area of the white surface, its position, and its reflectance. Accordingly, Stewart[6] measured the degrees black required in a black-white disk in a high illuminance to match a Gelb disk of 12-in. diameter with a small white surface either 0.5, 1.0, or 2.0-in. in diameter placed at the center and at various positions from center to the edge of the black disk. The reflectance of the white surface was not varied since Koffka had already found that the less the ratio of reflectances of the two surfaces, the less dark did the black disk appear. Stewart employed the binocular-septum matching technique in which the Gelb disk was viewed by the right eye against an unlighted black ground, while the matching disk, composed of white and black moveable sectors, was viewed by the left eye against a fully illuminated black ground.

The data plotted in Fig. 9 from Stewart show that the Gelb disk was blackest, i.e., required the largest amount of black in the matching disk, the nearer to the center and the larger the white surface was. The data plot so regularly on straight lines, except when the white surface is tangent to the black disk, that there can be no doubt as to the contrast origin of the Gelb phenomenon. Descriptive features of the Gelb phenomenon attest its contrast nature as well as the quantitative data. Observers reported that the darkening effect of the white surface was not uniform over the entire disk. The greatest darkening occurred in the vicinity of the small area and diminished with distance from it, as was expected from the contrast explanation. While there were individual differences among observers, the results for each were regular and consistent.

The extent of the triggering effect of the white surface in the Gelb phenomenon can be appreciated when it is realized that the ratios of the small white surfaces to the black disk were only 0.00017, 0.00068, and 0.00272 or, in terms of rounded percents, 0.02%, 0.07%, and 0.3%! Since the smallest white area gave considerable darkening of the Gelb disk, it is apparent that even the figure of 0.02% does not represent the limit of white area that may exert contrast effect under the conditions of the Gelb phenomenon.

The importance of the Gelb phenomenon is that it demonstrates very vividly that large contrast effect may be obtained with very small areas as well as with high luminances. The latter fact has been known and studied for a long time in investigations of glare from light sources. An effect similar to the Gelb effect is found if a small white patch is viewed in a high illuminance against a velvet-black background. In this case the white surface seems to glow, so great is the contrast effect under these conditions.

Theoretical Discussion

References to the von Bezold effect as a "mixture" or "spread" phenomenon imply that it is the result of irradiation or spread of excitation on the retina or something akin to a mosaic color mixture. While such mechanisms may be responsible for assimilation when line separations are no greater than can be resolved by the eye, they cannot explain assimilation with stimuli as widely separated as 57 ft, the widest separation at which some assimilation was reported in the study with Rohles (50-mm gray separation between lines viewed at a distance of 3 m). Similarly it is doubtful that tonic eye movements are an explanation for assimilation effects with stimuli subtending relatively large visual angles. On the other hand, if such explanations sufficed for assimilation they would still leave contrast unexplained. In view of our finding that assimilation and contrast vary along a single continuum in which there is a zone of neutrality, a single mechanism is preferable to handle all the facts. Such a mechanism is provided by the concept of retinal and/or central interaction.

Physiological models that envisage facts of facilitation, summation, and inhibition in spinal reflexes and in retinal processes based on spatial interaction between more or less distant neural areas have been proposed by a number of workers, among them Sherrington,[7] who drew a parallel between retinal function and spinal reflex activity. Polyak's histological studies have provided ample evidence for spatial interaction in the retina. In this connection he wrote: at least a "fraction of cone impulses [passes] to one or several horizontal cells from which it is distributed in all directions over the more or less distant rods and cones" by means of amacrine cells having lateral connections.[8] As pointed out by Ruch,[9] the convergence of many rods and cones on bipolar cells and of bipolar cells on ganglion cells provides a neural substrate for interaction of streams of impulses resulting in facilitation and inhibition in vision as well as in the case of spinal reflexes. In view of the opposed nature of assimilation and contrast, we must assume that if assimilation arises from summation or facilitation of impulses of a given kind or quantity, then contrast arises from inhibition of the same type of impulses in neighboring areas. We therefore assume that relatively equal impulses from neighboring areas summate, or at least do not inhibit one another, thereby giving rise to assimilation, while an intense excitation, arriving at some common synapse or neural pathway, inhibits a weaker one to give rise to contrast. Ruch also points out, "Convergence (provides)...the neural substrate for interaction of streams of impulses resulting in facilitation and inhibition phenomena" as well as the system of intraretinal association neurons.[9]

To explain the assimilation and contrast effects in these studies we need assume only that area acts like luminance, that is, increase in area has the same effect as increase in luminance, and that relatively weak interactions summate and intense stimulation in one area inhibits weaker stimulation in neighboring areas. Besides accounting for contrast and assimilation we can account on this basis for the neutral zone in which there is neither contrast nor assimilation as follows: if these two phenomena are mediated by a mechanism that results in summation with low, and inhibition with high, differentials of stimulation, then there should be neither summation nor inhibition at some intermediate differential in stimulation as is actually the case. The absolute levels at which there is neither assimilation nor contrast depend upon the luminances of the lines and of the intervening areas. One prediction from this theory has been indirectly verified: if the luminance of thin lines which give assimilation is increased sufficiently, then assimilation should give way to contrast. Inspection of Fig. 3 shows assimilation changes to contrast with constant linewidth as the ratio of gray width to line width <u>decreases</u> in accordance with the assumption that area and luminance function equivalently: the smaller gray area is equivalent to higher line-luminance with resultant contrast effect. Owing to the rapid fall-off of contrast with increasing distance from borders the area-luminance equivalence is probably not linear and needs to be determined independently.

The trigger type of contrast characteristic of the Gelb phenomenon requires only recognition of the fact that very small areas may, under certain conditions, exert very large contrast effects, or, in terms of the theory proposed here, a large differential in reflectance or luminance due to a very small area in the neighborhood of a large area may, under favorable conditions, inhibit excitations from the large area. More data are needed for a quantitative formulation of this theory of contrast and assimilation.

Acknowledgments

I am grateful to my co-workers, Dr. Frederick H. Rohles, Jr., and Dr. E. C. Stewart and Vernon Joy and Joe A. Steger, for their cooperation in the studies reported here and for permission to use previously published material so freely. In addition, I must express thanks to Dr. William Bevan for solving the problem of producing lines of all widths cleanly and quickly thus making possible inclusion of the results of the third study of assimilation and contrast in this report.

Table I. Mean judgments of the stimuli.[a]

Gray width (mm)	Linewidth (mm) 3	10	20	30	40	50	Total	Mean
3	2.73	5.59	6.31	6.40	6.68	6.86	34.57	5.76
10	3.41	4.81	5.54	5.86	6.45	6.31	32.38	5.40
20	4.09	5.40	5.68	5.90	6.13	6.40	33.60	5.60
30	3.91	5.09	5.31	5.81	6.09	6.72	32.93	5.49
40	4.36	4.90	5.04	5.72	5.95	6.13	32.10	5.35
50	4.59	4.77	5.59	5.31	6.00	6.27	32.53	5.42
Total	23.09	30.56	33.47	35.00	37.30	38.69	198.11	
Mean	3.85	5.10	5.58	5.83	6.22	6.45		

a From H. Helson and V. Joy, Psychol. Beiträge 6, 405 (1962).

Footnotes

1 W. von Bezold, Die Farbenlehre (Westerman, Braunschweig, Germany, 1874), The Theory of Color, American edition (Prang, Boston, 1876). The designs showing color assimilation have been reproduced in the following more accessible publications: R. M. Evans, An Introduction to Color (John Wiley & Sons, Inc., New York, 1948), pp. 181 and 192; R. W. Burnham, Am. J. Psychol. 66, 379 (1953).

2 A. Gelb, Handbuch norm. pathol. Physiol. 12, 594 (1930).

3 H. Helson and F. H. Rohles, Jr., Am. J. Psychol. 72, 530 (1959).

4 H. Helson and V. Joy, Psychol. Beiträge 6, 405 (1962).

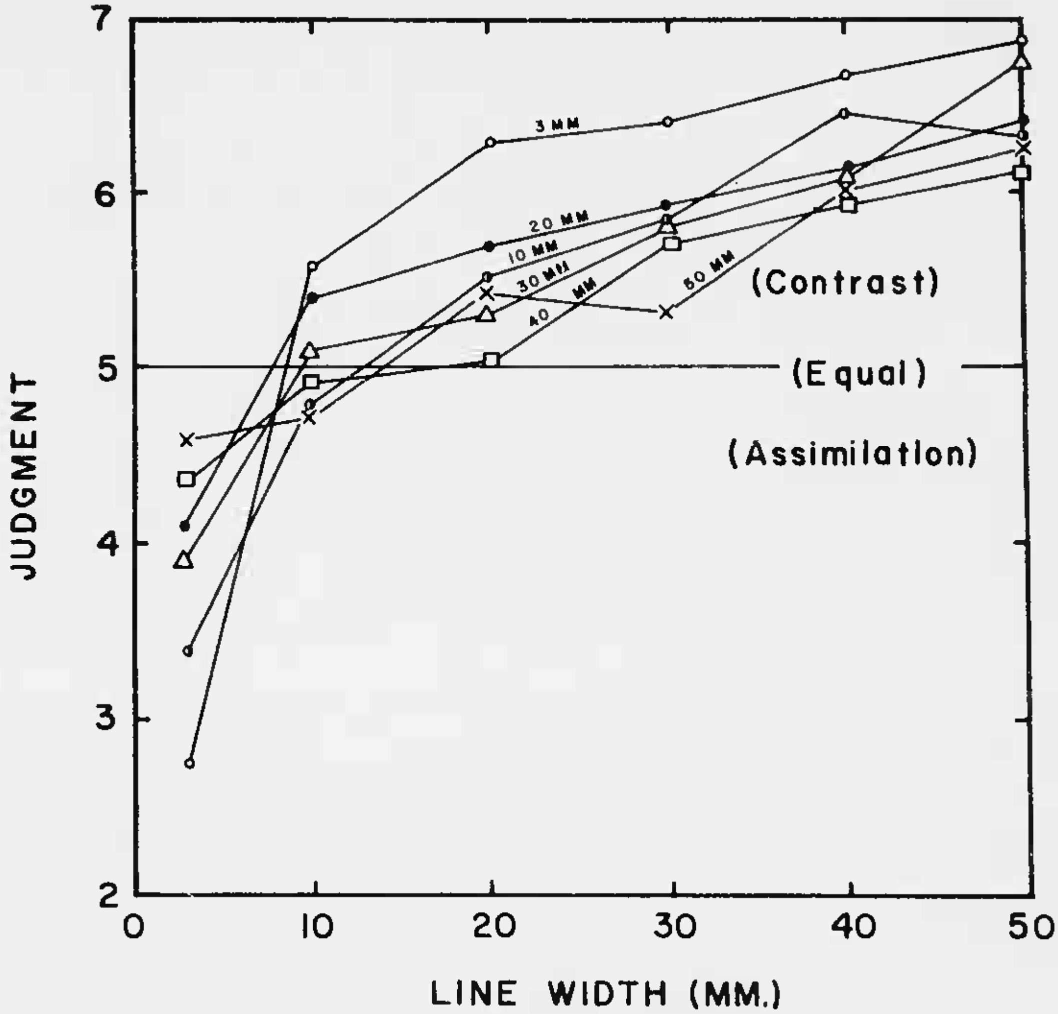
5 H. Helson and J. A. Steger (unpublished work).

6 E. C. Stewart, J. Exp. Psychol. 57, 235 (1959).

7 D. P. C. Lloyd in Howell's Textbook of Physiology, edited by J. F. Fulton (W. B. Saunders, Philadelphia, 1946).

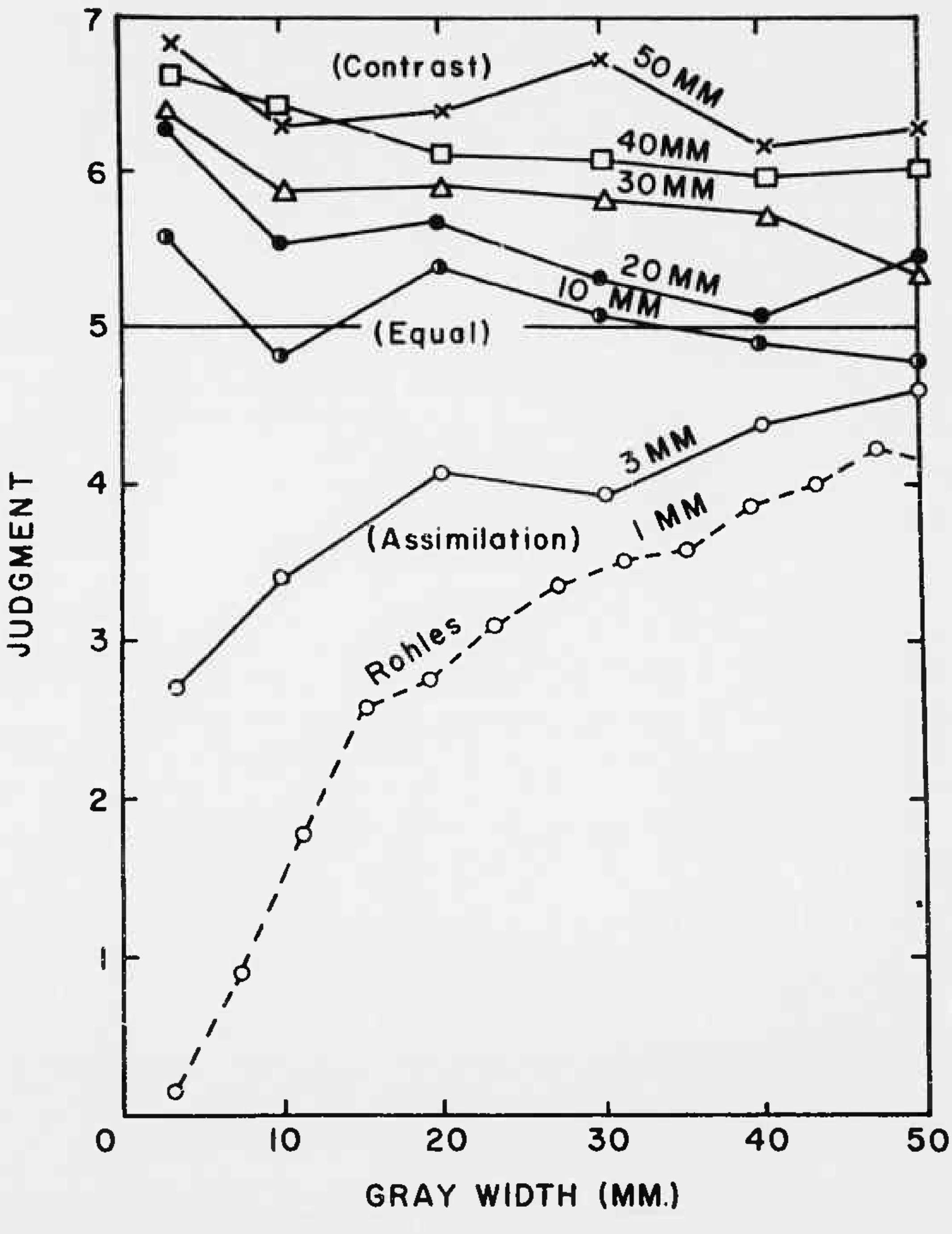
8 S. L. Polyak, The Retina (The University of Chicago Press, Chicago, 1941), p. 395.

9 T. C. Ruch in Howell's Textbook of Physiology (W. B. Saunders, Philadelphia, 1946), p. 476.



JUDGMENT
7
6
5
4
3
2
0
10
20
30
40
50
LINE WIDTH (MM.)
3 MM
20 MM
10 MM
30 MM
40 MM
50 MM
(Contrast)
(Equal)
(Assimilation)

Fig. 1. After Helson and Joy. Contrast and assimilation as a function of line width for each line separation.



7
6
5
4
3
2
1
0
10
20
30
40
50
JUDGMENT
GRAY WIDTH (MM.)
(Contrast)
50 MM
40MM
30MM
20 MM
10 MM
(Equal)
3 MM
1 MM
(Assimilation)
Rohles

Fig. 2. After Helson and Joy. Contrast and assimilation as a function of width of gray (line separation) for each width of line. The broken curve for 1-mm linewidth is from the study by Helson and Rohles.

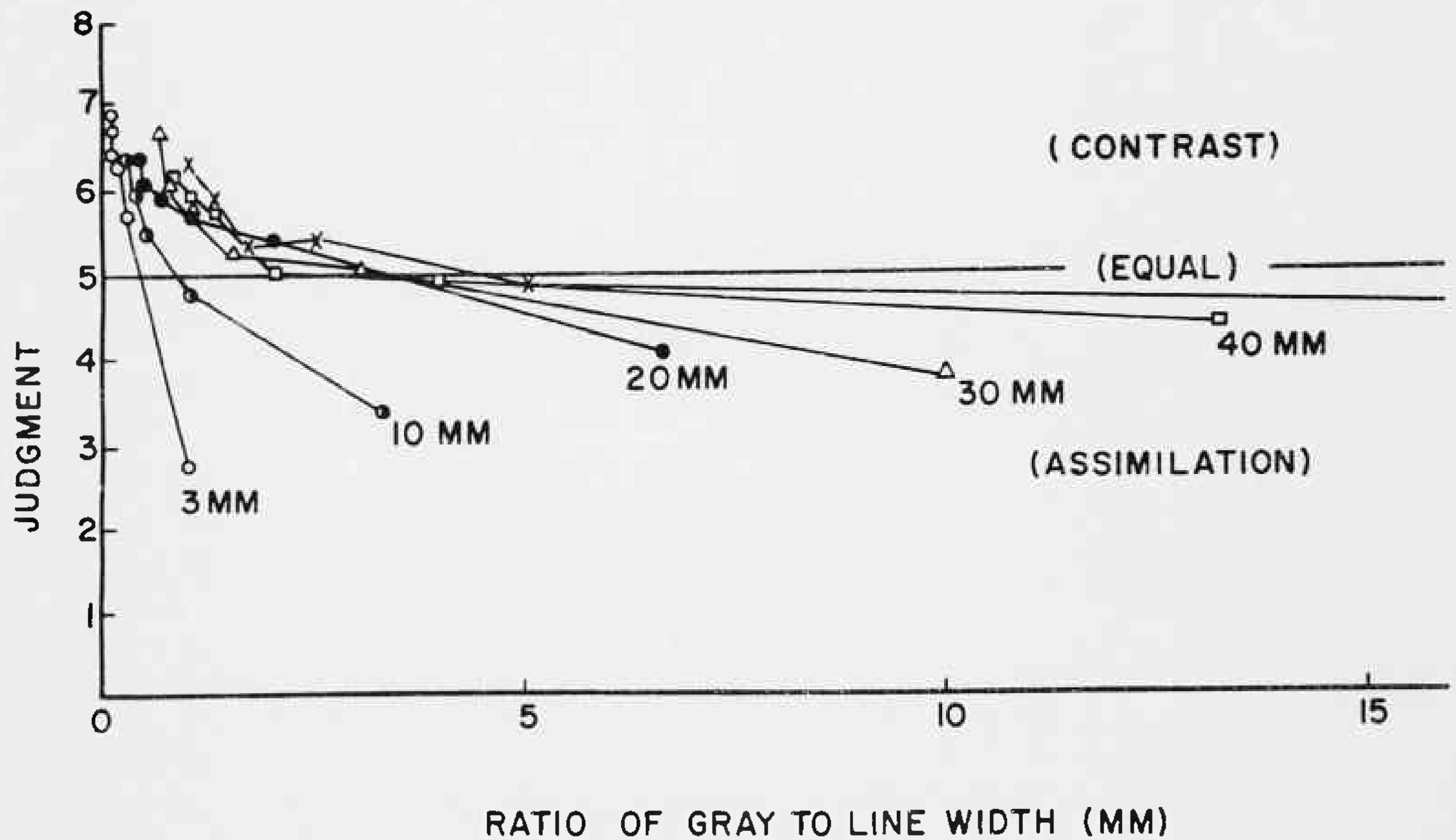
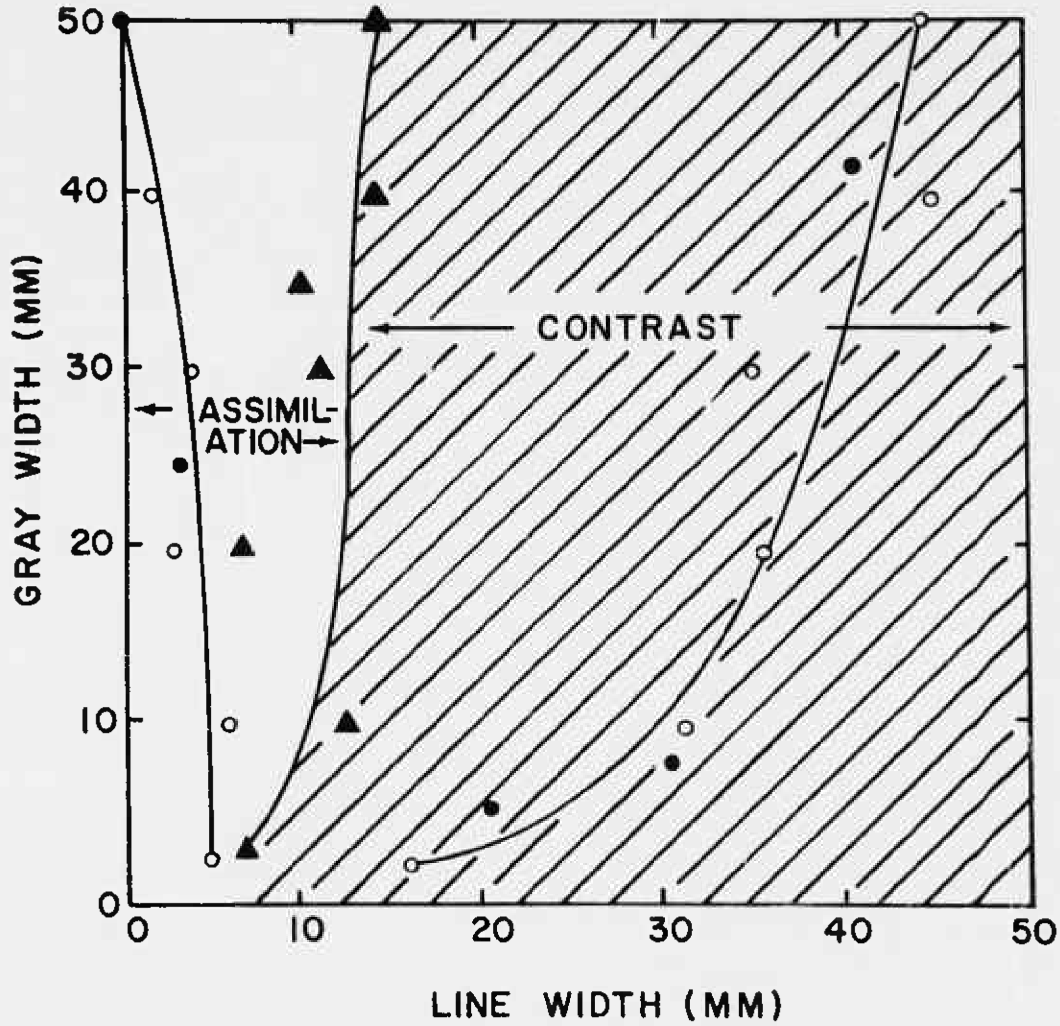
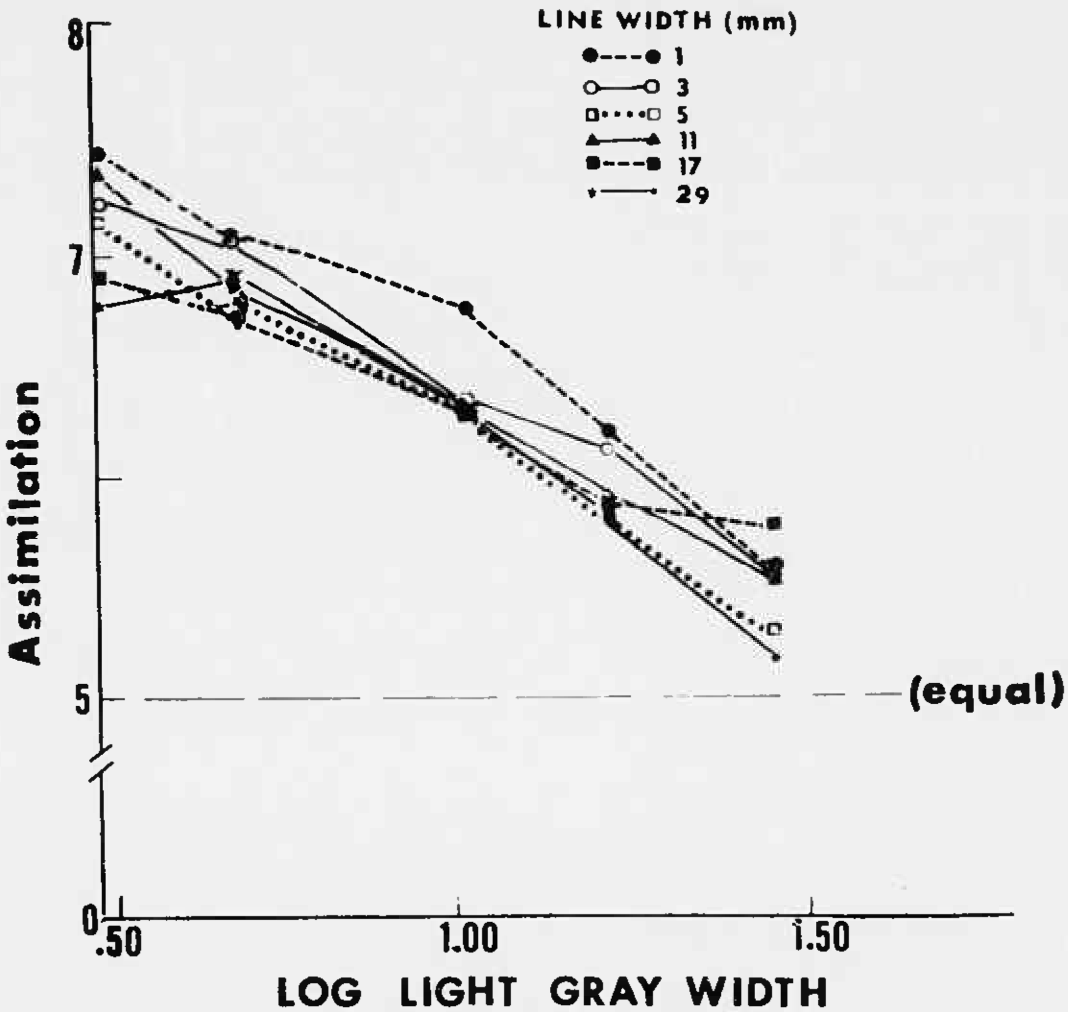

8
7
6
5
4
3
2
1
JUDGMENT
(CONTRAST)
(EQUAL)
(ASSIMILATION)
3 MM
10 MM
20 MM
30 MM
40 MM
0
5
10
15
RATIO OF GRAY TO LINE WIDTH (MM)

Fig. 3. After Helson and Joy. Assimilation and contrast as a function of the ratios of gray width to linewidth. Note the zone of equality where the curves cross the line of equals and the asymptotic approach of the assimilation points to the line of equals toward the right of the figure.



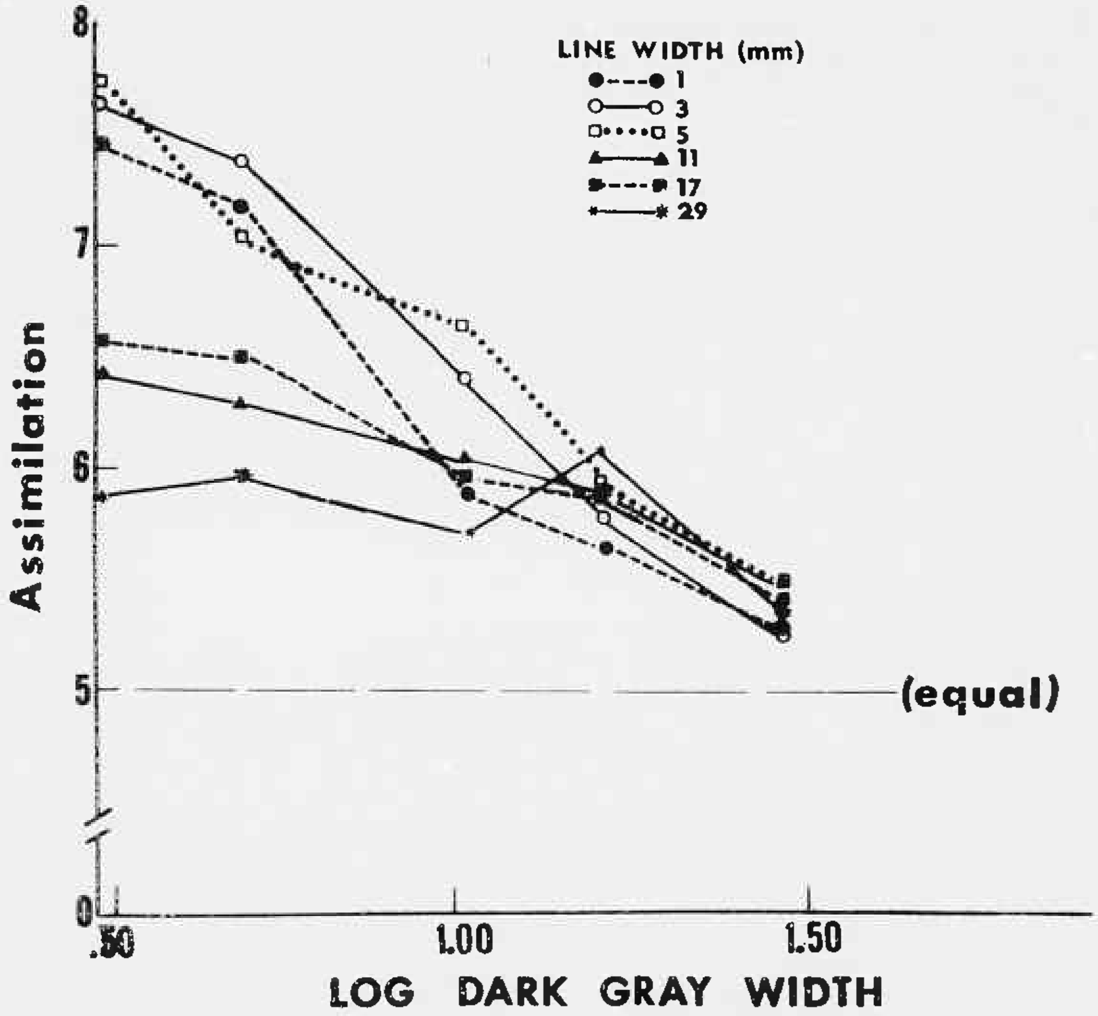
50
40
30
20
10
0
GRAY WIDTH (MM)
ASSIMIL-
ATION
CONTRAST
0
10
20
30
40
50
LINE WIDTH (MM)

Fig. 4. After Helson and Joy. Domains of assimilation and contrast. Each point in the plane represents the combination of linewidth and line separation that yields either assimilation or contrast.



LINE WIDTH (mm)
1
3
5
11
17
29
8
7
5
0
Assimilation
(equal)
.50
1.00
1.50
LOG LIGHT GRAY WIDTH

Fig. 5. Assimilation as a function of line separation (gray width) with background of 80% reflectance.



LINE WIDTH (mm)
1
3
5
11
17
29
Assimilation
8
7
6
5
0
(equal)
.50
1.00
1.50
LOG DARK GRAY WIDTH

Fig. 6. Assimilation as a function of line separation (gray width) with background of 14% reflectance.

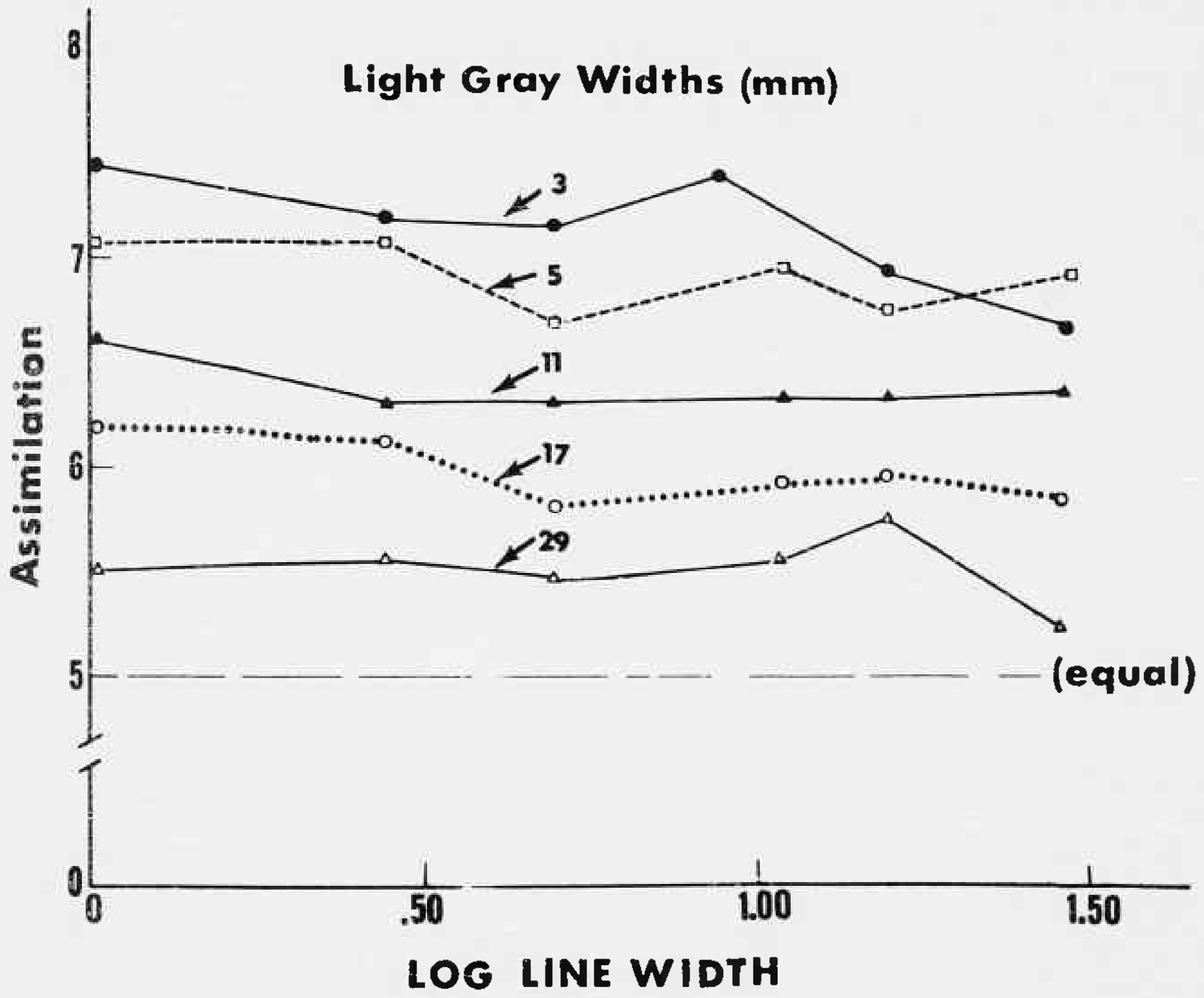

Light Gray Widths (mm)
3
5
11
17
29
(equal)
Assimilation
8
7
6
5
0
0
.50
1.00
1.50
LOG LINE WIDTH

Fig 7. Assimilation as a function of linewidth with hackground of 80% reflectance.

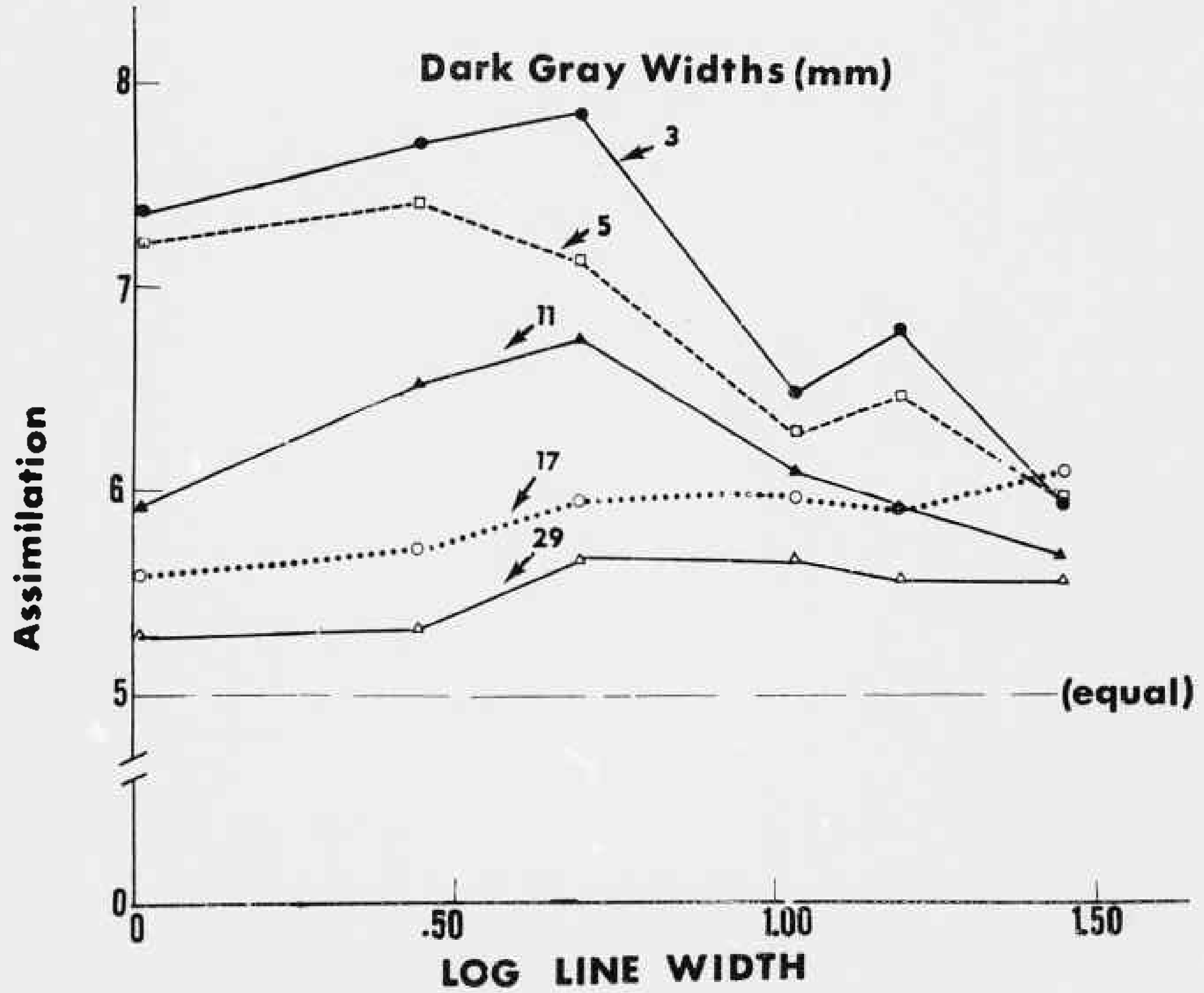

Dark Gray Widths (mm)
3
5
11
17
29
(equal)
Assimilation
8
7
6
5
0
0
.50
1.00
1.50
LOG LINE WIDTH

Fig. 8. Assimilation as a function of linewidth with background of 14% reflectance.

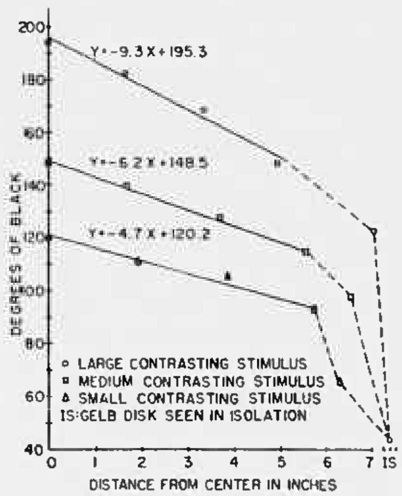

Y = -9.3 X + 195.3
Y = -6.2 X + 148.5
Y = -4.7 X + 120.2
DEGREES OF BLACK
200
180
160
140
120
100
80
60
40
0
1
2
3
4
5
6
7
IS
DISTANCE FROM CENTER IN INCHES
○ LARGE CONTRASTING STIMULUS
□ MEDIUM CONTRASTING STIMULUS
△ SMALL CONTRASTING STIMULUS
IS:GELB DISK SEEN IN ISOLATION

Fig. 9. After Stewart. Contrastive darkening of the Gelb disk as a function of distance from center of the white patch with size of patch the parameter in the individual curves.